HAL
open science

# Import des ARNt dans Plasmodium : sélection à l'entrée ?

Marta Cela Madinaveitia

## To cite this version:

Marta Cela Madinaveitia. Import des ARNt dans Plasmodium : sélection à l'entrée ?. Microbiologie et Parasitologie. Université de Strasbourg, 2018. Français. NNT : 2018STRAJ067 . tel-02292791v2

**HAL Id: tel-02292791**

**https://theses.hal.science/tel-02292791v2**

Submitted on 20 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# UNIVERSITÉ DE STRASBOURG

*ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE*

**UPR 9002**

# THÈSE présentée par :

**Marta CELA**

soutenue le : **20 Septembre 2008**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Sciences du vivant
Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

# Import des ARNt dans *Plasmodium :* Sélection à l'entrée ?

**THÈSE dirigée par :**

| | |
|---|---|
| **Mme. FRUGIER, Magali** | Directrice de recherche, Université de Strasbourg |

**RAPPORTEURS :**

| | |
|---|---|
| **Mme. ABERGEL, Chantal** | Directrice de recherche, Université de Marseille |
| **M. HELM, Mark** | Professor, Universität Mainz |

---

**AUTRES MEMBRES DU JURY :**

| | |
|---|---|
| **M. GIEGE, Philippe** | Directeur de recherche, Université de Strasbourg |

Je voudrais remercier toutes les personnes qui ont contribué d'une façon ou d'une autre à la réalisation de cette thèse.

Je tenais à remercier mes membres du jury, Chantal Abergel, Mark Helm et Philippe Giegé, qui me font l'honneur d'évaluer mon travail de thèse. Je remercie le CNRS, le Labex NetRNA et la FRM pour rendre possible la réalisation de ce travail. Merci aussi au Professeur Eric Westhof de m'avoir accueilli au sein de son unité, et à Pascale Rombi de rester toujours accessible et à l'écoute.

Magali, je voulais te remercier cette opportunité. Cette thèse m'a aidé à évoluer dans le domaine professionnel, mais aussi dans le personnel. Je remercie tous tes conseils, tes corrections, et tous les bons moments passés au labo. Tu as cru en moi, tu m'as donné confiance, et tu m'as poussé à m'améliorer. Merci pour ces 4 ans d'intéractions (hé oui, je devais le caser quelque part) humaines et professionnelles. Je garderais toujours un bon souvenir de mes premiers pas dans la science.

Merci à toute l'équipe, Anne, Caro, Joëlle, Claude et Bernard. Chacun contribue à sa façon à la bonne ambiance du labo. Merci Anne et Joëlle pour nos petites discussions. Merci de votre soutient, qui m'a énormément aidé. Joëlle, cette complicité entre nous deux va vraiment me manquer. Et je sais que ça va beaucoup te manquer de me taquiner ;). Je remercie ton aide précieuse, ton support, ta bonne humeur et tes conseils. Caro, sans toi, le labo ne serait pas le même. Toujours discrète, mais toujours là, la première à râler quant c'est nécessaire XP. Merci. Un grand merci à tous.

Mais bien sur que je ne vous ai pas oublié ! Votre support, surtout ces derniers mois, a été inestimable. Merci a mes trois « chicos ». Vous avez pris soin de moi, avec ces pauses glace, et ces soirées folles. Merci Raph, José et Cédric. On se reverra pour l'enterrement de vie de jeune garçon de José ;)

J'en profite pour remercier Justine ! C'est grâce à toi, qui m'as accueilli dans le master, que j'ai vu que les français pouvaient être sympas XP. Merci mi amor pour toutes ces années d'amitié, toutes ces soirées bouffe, qui m'ont aidé a déconnecter du labo.

Je veux aussi remercier Jana et Delphine, qui m'ont accueilli dans le labo, et au sein de la petite famille 443. Merci maman, et merci grande petite sœur. Jamais je n'oublierai notre complicité, nos fous rires tard au labo, et nos réunions d'équipe/soirées filles. Avec toi, Delphine, j'ai partagé des moments auxquels je ne pensais pas survivre, comme les après midis avec ta playlist à 8 chansons (répétées à l'infini), ou les looooongues heures devant le microscope à compter des points verts et rouges. Merci à toutes les deux.

Et comment aurais-je pu survivre à la rédaction sans la team doctorands ? Merci Alexi, Damien et Elena de partager ces moments avec moi. Vous savoir dans ma même situation m'a donné la force pour continuer et finir ce manuscrit. Et merci Benjamin pour tes conseils, et pour mes nouveaux amis dessinés sur ma paillasse !

Annika, ma petite stagière, je n'oublierais jamais les mois passés ensembles. Grâce à toi j'ai perdu la peur de l'enseignement (juste un peu), et j'ai plus de confiance en moi. Même malgré la lutte contre ton entêtement, je garde un très bon souvenir de ton stage XP

Renaud, un grand merci de m'avoir accueilli dans ton labo, et pour toute la liberté que tu m'as donnée. Ces semaines à Charleston étaient très amusantes et très importantes pour moi. Je te remercie aussi tes questions pour réfléchir (« et qu'est ce que tu vaux faire après la thèse ? »). Je sais que tu poses ces questions à tous tes étudiants parce que tu aimes les mettre dans le pétrin, mais tu restes toujours ouvert à la discussion, prêt pour aider tes petits étudiants perdus. Je tiens à remercier également Ariel, Evelyn et Jason pour leur accueil dans le laboratoire (Thank you guys !), et Stéphanie pour toute son aide.

L'équipe moustique a également constitué une partie importante de ma thèse. J'ai beaucoup apprécié les matins à l'insectarium avec Steph, Amandine, Hélène et Nath, et les petits mots surprise en espagnol d'Eric. Merci aussi à tout le labo (Alice-Ane, Mélanie, Vru, Brice, Emily, Roenik, Katharina, Dennis, Moustiquette, Julien, Randy,…) pour la superbe ambiance. Cyril, cette soirée laser game était une idée magnifique ! Noémie, tu sais que je t'aime, mais tu es une cible facile ;)

Merci David et Karine pour votre inestimable aide avec les souris. Et merci aussi pour ces longues discussions avant d'aller au laboratoire. Malgré les difficultés, vous allez toujours de l'avant avec le sourire, et c'est quelque chose que j'admire chez vous.

Je sais qu'il y a beaucoup de personnes que je n'ai pas mentionnées, qui ont été importantes pour moi d'une manière ou d'une autre au cours de ces quatre années, comme Isabelle, Alain, Mélanie, Florian, Fafa, Paula, Marine, Safi… Ne pensez pas que je vous ai oublié. Je garde beaucoup de souvenirs de ces années à Strasbourg.

En el lado español de mi tesis, no puede faltar mi dream team. Gracias chicos por esas barbacoas, esos momentos de risas, de lloros, de amor, de odio. Gracias por estar ahí. Esto no habría sido lo mismo sin vosotros, Simone, Nico, y sobre todo sin mis churris Rata y Sandrius. Chicas, espero que lloréis el día de mi tesis todo lo que estoy llorando yo ahora.

Hay otra persona que ha sido muy importante durante estos años. Me ha ayudado cuando me he sentido lejos, cuando me he sentido sola, y cuando me he sentido incomprendida. Muchas gracias, mi Señora del Caos. Y gracias también al Señor del Caos, por cuidar de mi Vane.

En cuanto a mis biologuillos, si que es verdad que nos hemos distanciado. Me he venido muy lejos, habéis seguido con vuestras vidas, y no siempre lo he llevado bien. Aún así, habéis estado ahí, esperando al acecho que os dijera los días que estaba en Madrid para quedar conmigo y darme ánimos.

Quiero terminar dando las gracias a mi familia. A mi Panda, que me ha hecho reír en mis momentos malos. Que ha venido a verme (o a ver los Ponts Couverts más bien) cada vez que ha podido. Que siempre está ahí, como mis padres y Carlos. Quiero agradecer a mis padres su paciencia infinita, sobre todo, durante mi carrera, también infinita. ¡Al final no he

salido mal! Muchas gracias por las llamadas diarias para ver qué tal estoy, por uberizarme, y por animarme y apoyarme en la distancia. Y gracias a Carlos, uno de los que más me ha sufrido durante esta tesis. Gracias por todos esos momentos, buenos y malos. Gracias por venir a verme tan a menudo, por preocuparte por mi, por apoyarme y darme ánimos.

Gracias a todos.
Merci à tous.

# SOMMAIRE

INTRODUCTION 1

I. PLASMODIUM 3

1. Brève présentation du paludisme 3

2. Le cycle de vie de Plasmodium (Figure 2) 6

3. Les différents génomes de Plasmodium 9

*3.1. Génome nucléaire 9*

*3.2. Génome mitochondrial 10*

*3.3. Génome de l'apicoplaste 10*

4. L'apicoplaste 11

5. Interaction hôte/parasite : échanges moléculaires 12

*5.1. Endocytose de l'hémoglobine 13*

*5.2. Mise en place de la machinerie d'import des nutriments dans le globule rouge infecté 14*

II. LES ARN DE TRANSFERT 17

1. Structure des ARNt 17

2. Maturation des ARNt 19

3. Modifications post-transcriptionnelles des ARNt 21

*3.1. Modifications dans les boucles D et T 23*

*3.2. Modifications dans la boucle Anticodon 23*

*3.3. Les modifications dans le reste de la molécule 23*

4. Aminoacylation des ARNt 24

*4.1. Les aminoacyl-ARNt synthétases 24*

*4.2. Spécificité de la réaction d'aminoacylation 26*

III. CIRCULATION DES ARNT 29

1. Circulation des ARNt entre le noyau et le cytoplasme 29

*1.1. Transport de l'ARNt du noyau vers le cytoplasme 30*

*1.2. Transport du cytosol vers le noyau 31*

**2. Circulation des ARNt du cytoplasme vers la mitochondrie** **32**

***2.1. Mécanisme d'import des ARNt dans la mitochondrie de Saccharomyces cerevisiae*** ***33***

***2.2. Mécanisme d'import des ARNt dans la mitochondrie des Kinetoplastides*** ***34***

***2.3. Mécanisme d'import des ARNt dans la mitochondrie de plantes*** ***35***

**IV. LES ARNT DE PLASMODIUM** **39**

**1. Rapide survol des caractéristiques structurales des ARNt de *Plasmodium*** ***39***

***1.1. Les ARNt de l'apicoplaste*** ***39***

***1.2. Le cas particulier de la mitochondrie*** ***40***

***1.3. Les ARNt cytosoliques*** ***40***

***1.4. Les ARNt et le repliement cotraductionnel des protéines*** ***41***

**2. tRip et l'import des ARNt exogène dans *Plasmodium*** **42**

***3. Les domaines OB-fold dans les pathogènes unicellulaires eucaryotes*** ***44***

**V. OBJECTIFS DE MA THESE** **57**

**1. La sélectivité de tRip *in vitro*** **57**

**2. Présence des ARNt de l'hôte dans le parasite au stade sanguin** **57**

**3. La tyrosylation dans l'apicoplaste de *Plasmodium falciparum*** **58**

**RESULTATS** **59**

**I. IDENTIFICATION DES ELEMENTS D'IDENTITE DE L'ARNT$^{TYR}$ APICOPLASTIQUE** **61**

**II. LA SELECTIVITE DE TRIP *IN VITRO*** **93**

**1. Le MIST** **93**

***1.1. Principe du MIST*** ***93***

***1.2. Purification de la protéine PftRip*** ***93***

***1.3. Purification de l'ARNt total de cellules HeLa*** ***96***

***1.4. MIST*** ***98***

*a. Isolement des complexes ARNt/PftRip* *98*

*b. Les puces à ARNt* *99*

*c. Quantification et interprétation des résultats* *100*

***1.5. Etude de l'affinité de PftRip pour différents transcrits d'ARNt* in vitro** ***103***

***1.6. Rôle des modifications post-transcriptionnelles dans la discrimination des ARNt par PftRip*** ***108***

***1.7. Recherche d'un profil de modifications post transcriptionnelles modulant l'affinité de PftRip pour les ARNt*** ***113***

***1.8. Conséquences sur la biologie du parasite*** ***116***

**2. Le SELEX** **118**

***2.1. Préparation du matériel*** ***119***

***2.2. Les cycles de sélection*** ***121***

***2.3. A la recherche d'un motif commun*** ***124***

***2.4. Comparaison des affinités des aptamères*** ***125***

***2.5. Caractérisation des aptamères 23 et 26*** ***128***

*a. Affinité pour PftRip* *128*

*b. Structures en solution* *132*

***2.6. Conclusions*** ***132***

**III. PRESENCE DES ARNT DE L'HOTE DANS LE PARASITE AU STADE SANGUIN** **135**

**1. Y a t-il de l'ARNt dans les globules rouges ?** **136**

**2. Mise au point d'un protocole de purification des ARN de parasite** **137**

***2.1. Effet de la RNase*** ***138***

***2.2. Rendements*** ***138***

***2.3. Qualité des échantillons*** ***139***

**3. Peut-on éviter l'utilisation de RNases ?** **141**

**4. Peut-on détecter des ARNt de souris parmi les ARNt du parasite ?** **142**

**5. Conclusions** **143**

**CONCLUSION ET PERSPECTIVES** **157**

**MATERIEL ET METHODES** **167**

**I. MATERIEL** **169**

**1. Produits chimiques** **169**

**2. Nucléotides, oligonucléotides et acides aminés** **170**

**3. Acides nucléiques non commerciaux** **171**

**4. Enzymes de restriction et de modification** **171**

**5. Enzymes préparées au laboratoire** **171**

**6. Plasmides** **172**

**7. Cellules pour clonage et production d'ARNt** **172**

**II. METHODES** **173**

**1. Préparation des parasites** **173**

***1.1. Souches de* Plasmodium *utilisées dans cette étude*** ***173***

***1.2. Infection des souris*** ***173***

*a. Inoculation des parasites* *174*

*b. Suivi de l'infection* *175*

*c. Prélèvement du sang parasité par ponction cardiaque* *175*

*d. Stocks de parasite au stade sanguin* *175*

***1.3. Les moustiques*** ***176***

*a. Maintient de la lignée* *176*

*b. Infection des moustiques* *177*

*c. Dissection des glandes salivaires* *177*

**2. Préparation des protéines recombinantes** **178**

***2.1. PftRip (tRNA import protein de* Plasmodium falciparum*)*** ***178***

*a. Protocole de purification de PftRip* 179

*b. Protocole de purification des deux formes du domaine C-Terminal de PftRip* 180

***2.2. Tyrosyl-ARNt synthétase apicoplastique de* Plasmodium falciparum *: PfapiTyrRS*** ***181***

**3. Préparation des acides nucléiques** **184**

***3.1. Purification d'ARNt natifs à partir de cellules HeLa*** ***184***

*a. Préparation de l'ARNt total* 184

*b. Préparation de la colonne d'affinité* 184

*c. Hybridation* 185

*d. Elution* 185

*e. Analyse par spectrométrie de masse* 185

***3.2. Extraction d'ARN total de* Plasmodium berghei** ***186***

*a. Purification des globules rouges parasités* 186

*b. Lyse globules rouges parasités* 186

*c. Elimination des ARN contaminants (de l'hôte)* 187

*d. Extraction de l'ARN* 187

***3.3. Production de transcrits d'ARNt* in vitro** ***188***

*a. Conception des cassettes de clonage* 188

*b. Transcription* 189

***3.4. Renaturation*** ***190***

**4. Etudes structurales et fonctionnelles des ARN** **190**

***4.1. Marquages radioactifs des acides nucléiques*** ***190***

*a. Marquage en 5'* 190

*b. Marquage en 3' des ARNt* 191

***4.2. Etudes du Repliement en solution des ARNt*** ***191***

*a. Coupure au plomb* 191

*b. Utilisation de sondes enzymatiques* 192

***4.3. Etudes fonctionnelles*** ***192***

*a. Northern blots* 192

*b. Intéractions ARN/protéines* 193

***4.4. Aminoacylation des ARNt*** ***194***

*a. Activation de l'acide aminé* *194*

*b. Aminoacylation des ARNt* *194*

**5. Techniques de sélection *in vitro*** **195**

***5.1. MIST*** ***195***

*a. Le principe* *195*

*b. Conception de la puce à ARNt* *197*

*c. Préparation des échantillons d'ARNt compléxés ou pas à tRip* *197*

*d. L'analyse des donnés* *198*

***5.2. SELEX*** ***198***

*a. Le principe* *198*

*b. Construction de la bibliothèque d'aptamères* *199*

*c. Cycles de sélection* *200*

*d. Clonage des candidats* *202*

**REFERENCES** **203**

# INTRODUCTION

## I. PLASMODIUM

### 1. Brève présentation du paludisme

Le paludisme est une maladie endémique des pays tropicaux, très répandue de nos jours (Figure 1). En 2016, cette maladie a été à l'origine de 445000 morts. La zone du monde la plus touchée est l'Afrique. Sur les 216 millions de cas signalés en 2016, 90 % sont survenus dans cette région, suivis de loin par la région d'Asie du Sud-Est (7 %) et la Méditerranée orientale (2 %). Malgré la tendance générale à la rémission (les cas de paludisme ont diminué de 18 % entre 2010 et 2016), il semble que le paludisme resurgisse dans les zones tropicales d'Amérique et de la Méditerranée orientale (OMS, 2017).

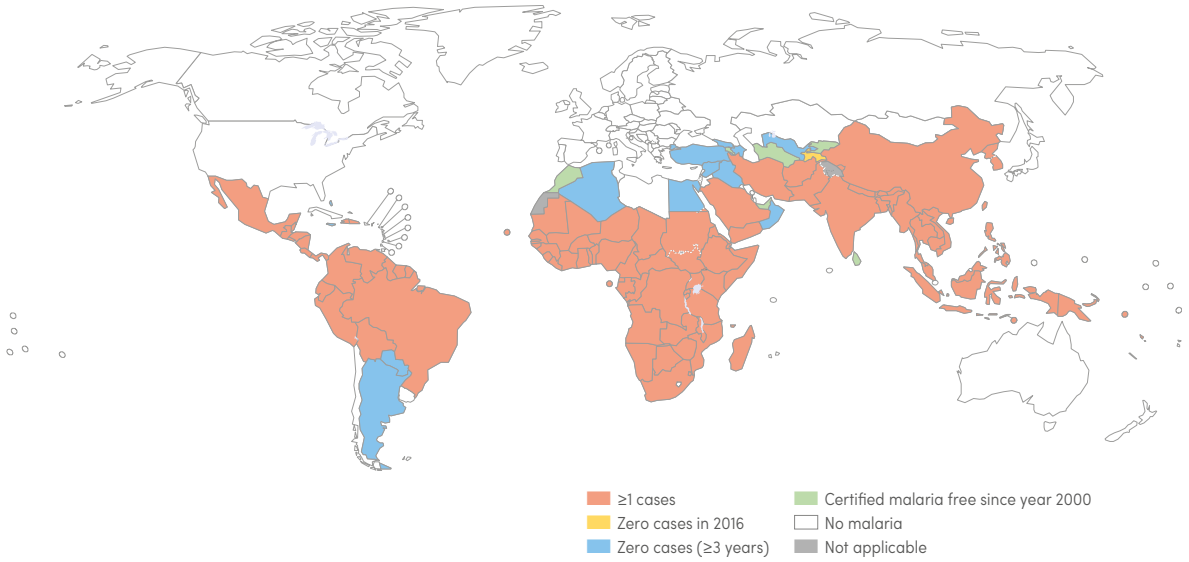



**Figure 1 : Répartition du paludisme en 2016 (OMS, 2017)**

Cette maladie est causée par un protozoaire du genre *Plasmodium*, appartenant à la famille des *Apicomplexa*. Ce parasite intracellulaire infecte deux hôtes lors de son cycle biologique: un hôte vertébré, et un moustique (*Anopheles*), qui va transmettre la maladie (Figure 2). L'hôte vertébré peut être un oiseau, un reptile ou un mammifère. Par exemple, *Plasmodium berghei*, se développe spécifiquement chez la souris. Seules cinq espèces peuvent provoquer le paludisme chez l'Homme : *P. falciparum*, *P. malariae*, *P. ovale*, *P. vivax* et *P. knowlesii*, ces trois dernières pouvant aussi infecter certains singes. *P. falciparum* est le parasite le plus répandu en Afrique et le plus dangereux. Il est responsable de 99% des cas

déclarés dans cette zone en 2016. Cependant, *P. vivax* prédomine dans la région des Amériques (64 % des cas), et est responsable de 4% des cas en Asie du Sud-Est (OMS, 2017).

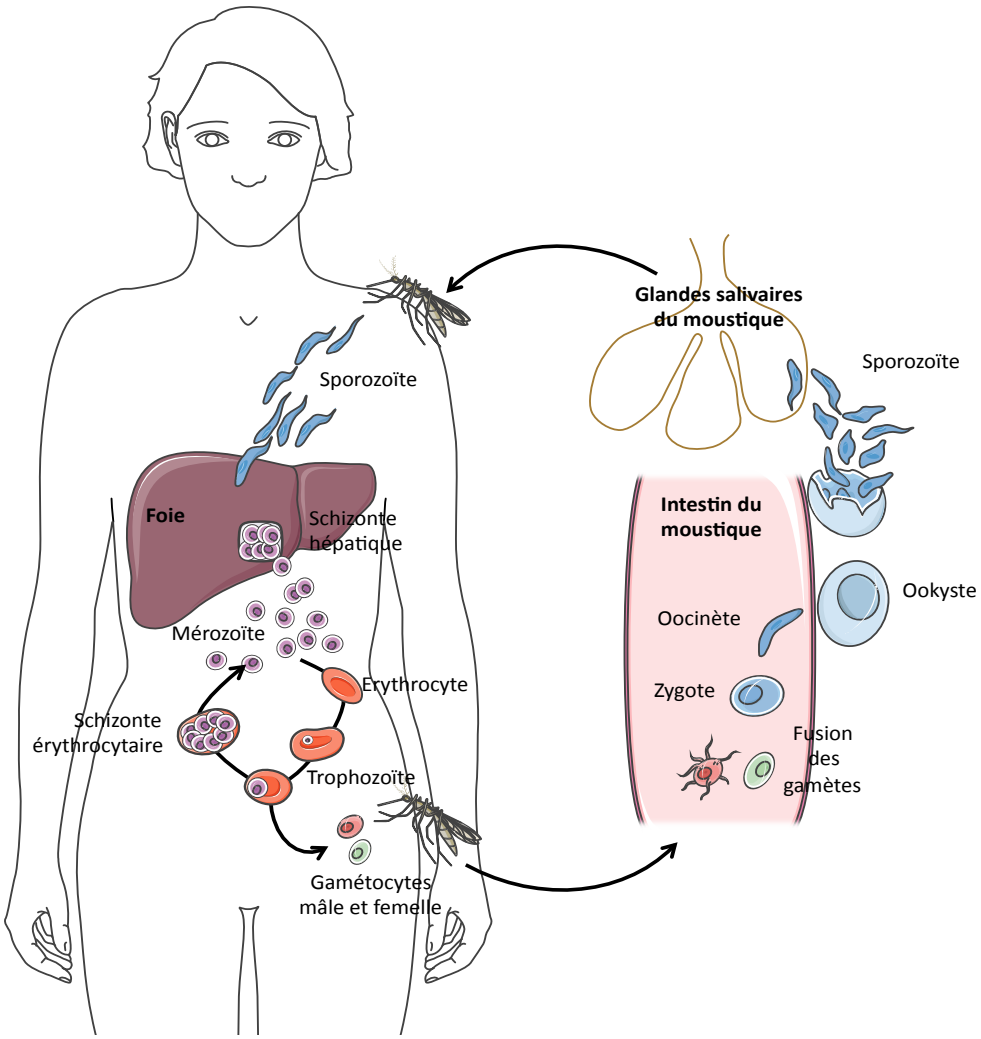



**Figure 2 : Cycle de vie de *Plasmodium falciparum*.** Le développement de *Plasmodium* se partage entre son vecteur l'anophèle femelle (cycle de reproduction sexuée), à droite, et l'Homme (reproduction asexuée), à gauche. Chez l'Homme, le cycle de vie comporte deux grandes étapes : le cycle hépatique et le cycle érythrocytaire. Chaque stade de développement de *Plasmodium* est indiqué et est expliqué dans le texte.

Les symptômes apparaissent entre 12 et 14 jours après la piqûre d'un moustique femelle infectée, coïncidant avec le développement du parasite dans le sang. Les symptômes initiaux (malaise, mal de tête, fatigue, nausée) sont très généraux. Le plus caractéristique est l'accès palustre, c'est à dire des cycles de forte fièvre récurrents. Cependant, la maladie peut causer des complications, ce qui provoque des symptômes tels que collapsus circulatoire,

détresse respiratoire aiguë, une anémie sévère, l'insuffisance d'un organe, et dans le cas du paludisme cérébral, un coma et/ou des convulsions (Marks, *et al.*, 2014). Avec un diagnostic précoce et des soins adéquats, cette maladie ne présente pas un risque de décès. Cependant, les mécanismes qui déclenchent ces complications ne sont pas connus, ce qui explique la nécessité d'un diagnostic précoce pour minimiser le risque de développer un paludisme grave.

Le traitement contre cette maladie est simple et efficace. Il est composé d'une combinaison de deux médicaments : un dérivé de l'artémisinine et un antipaludéen de longue persistance, tel que la méfloquine ou la pipéraquine. Leur effet contre *Plasmodium* perdure entre 4 et 6 semaines. Ce traitement, également appelé ACT (Artemisin-based Combination Therapy), découle de la nécessité de combiner deux antipaludéens pour palier à l'apparition des résistances aux médicaments.

Environ 30 ans après le développement du premier antipaludéen synthétique, la chloroquine, dans les années 1930, des formes de résistance apparaissent. Il a fallu seulement 10 ans pour que *P. falciparum* soit résistant dans 90 % des cas à cette molécule (Packard, 2014). La raison principale de combiner deux antipaludéens est d'éviter l'émergence de telles résistances. Cependant une résistance à l'artémisinine a déjà été observée en Asie du Sud-Est.

A l'heure actuelle, aucun vaccin efficace n'a été développé, malgré la grande implication de la communauté scientifique. Le vaccin appelé RTS, S/AS01 est le seul vaccin qui a été évalué à grande échelle dans la population (dans sept pays d'Afrique). Cependant, il ne protège que 51% des enfants pendant la première année après la vaccination, et seulement 26% trois ans plus tard (OMS, 2016). La difficulté de développer un vaccin efficace contre ce parasite réside dans la complexité de son cycle biologique (Figure 2). A chaque étape, l'expression des protéines, et donc des antigènes, varie, facilitant ainsi l'évasion du parasite du système immunitaire de l'hôte.

Le meilleur moyen de protection contre le paludisme est donc la prévention. Des mesures pour lutter contre le vecteur, comme l'utilisation de moustiquaires imprégnées d'insecticide, ont été mises en place dans les zones endémiques. Cependant, les moustiques commencent à développer eux aussi des résistances aux insecticides.

L'apparition de résistances, à la fois dans le microorganisme et dans le vecteur, et l'absence d'un vaccin efficace, constituent un sérieux problème pour le contrôle de la maladie.

C’est pour ça que l’étude de *Plasmodium* est essentielle, pour identifier de nouvelles cibles et concevoir de nouveaux médicaments antipaludiques.

## 2. Le cycle de vie de *Plasmodium* (Figure 2)

Le cycle de vie de *Plasmodium* implique deux hôtes. La femelle du moustique *Anopheles* (hôte invertébrée) est responsable de la dissémination du parasite et infecte l'hôte vertébré (oiseaux, reptiles ou mammifères). A chaque étape du cycle, *Plasmodium* subit des changements morphologiques et métaboliques, ce qui le fait s'adapter au tissu dans lequel se trouve. Les différents stades de *Plasmodium* peuvent être divisés en deux types : infectieux et prolifératif.

Le cycle commence lorsqu'un sporozoïte, forme infectieuse retrouvée dans les glandes salivaires du moustique, atteint son hôte vertébré lors du repas sanguin d’un moustique. Quand le moustique pique le mammifère, il injecte sa salive contaminée avec des sporozoïtes. Une dizaine de sporozoïtes suffisent à infecter un hôte vertébré (Mota et Rodriguez, 2002). Deux mécanismes d'entrée des sporozoïtes dans les cellules ont été décrits (Figure 3).

- Le premier mécanisme implique l'invagination de la membrane de la cellule hôte (Figure 3C). Dans ce cas, l’extrémité apicale du sporozoïte forme une « moving junction » (une jonction intime) avec la membrane de la cellule à envahir, et exerce une pression pour forcer la formation d’une vacuole (Amino, *et al.*, 2008). Une fois à l’intérieur, le parasite remodèle la membrane de la vacuole parasitophore pour se mettre à l’abri de la dégradation lysosomale et s’y multiplier. Cette étape est essentielle pour la prolifération du parasite (Amino, *et al.*, 2008, Risco-Castillo, *et al*., 2015).

- Dans le deuxième mécanisme, les sporozoïtes ne s’installent pas, ils déchirent la membrane et traversent la cellule hôte (Figure 3A). Risco-Castillo et coll. (2015) proposent dans leur article que certains sporozoïtes forment d’abord une vacuole de transition, et dans un deuxième temps, ils s’en libèrent grâce à une protéine similaire à une perforine (Perforine-Like Protein 1, ou PLP1) (Figure 3B). De même, Yang *et al.* (2017) et Ishino, *et al.* (2004) confirment que PLP1 et SPECT (Sporozoite microneme Protein Essential for Cell Traversal),

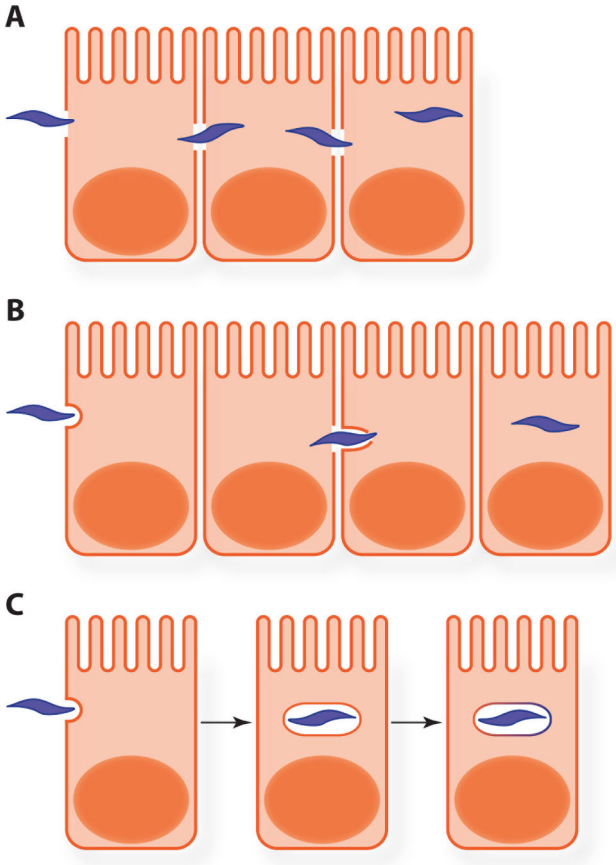



**Figure 3 : Différentes stratégies de *Plasmodium* pour envahir les cellules hôtes.** La traversée cellulaire **(A)** et **(B)** précède l'étape d'invasion et multiplication du parasite **(C)**. **(A)** Une première stratégie est de traverser directement la cellule hôte en formant une brèche dans la membrane. La brèche cicatrise après le passage du sporozoïte. **(B)** Une deuxième stratégie consiste en la formation d'une vacuole transitoire, le sporozoïte s'en évade et se retrouve dans le cytoplasme de la cellule hôte. **(C)** Le parasite entre dans la cellule hôte en formant une vacuole parasitophore, qu'il va transformer pour échapper à la dégradation et pour s'y multiplier.

des protéines localisées dans les micronèmes (Figure 4), sont essentielles pour la traversée cellulaire dans le foie. La traversée cellulaire est très rapide. Elle a été décrite lors de la migration du parasite dans la peau vers le système circulatoire, et quand le parasite quitte le flux sanguin pour infecter le foie. D'une part, la traversée cellulaire permet au parasite de quitter rapidement un environnement hostile (Menard, *et al*., 2013). D'autre part, elle permet aussi au parasite de leurrer le système immunitaire de l'hôte. En effet, le parasite passe à travers les cellules de Küpffer dans le foie (Tavares, *et al*., 2013, Kori, *et al*., 2018). Ces cellules sont des macrophages responsables de la phagocytose des pathogènes qui atteignent le foie par le sang (Dixon, *et al*., 2013). En traversant les cellules de Küpffer, le parasite force les cellules à se réparer, évitant la phagocytose. De plus, le parasite modifie les cellules qu'il

traverse. Ainsi, après le passage du sporozoïte, certaines protéines du parasite subsistent dans le cytoplasme de la cellule hôte, notamment, la protéine circumpsporozoïte (CSP), un composant majeur de la surface de ce stade du parasite. Les cellules traversées présentent la CSP à leur surface comme antigène. Elles sont alors reconnues par le système immunitaire, ce qui crée de fausses cibles, et facilite l'évasion des parasites (Mota et Rodriguez, 2001, Menard, *et al*., 2013, Tavares, *et al*., 2013, Kori, *et al*., 2018). Enfin, une controverse existe dans certaines publications qui proposent la traversée cellulaire comme nécessaire pour l'activation, l'invasion et la multiplication du sporozoïte dans un hépatocyte (Mota et Rodriguez, 2001, Mota et Rodriguez, 2002, Kori, *et al*., 2018).

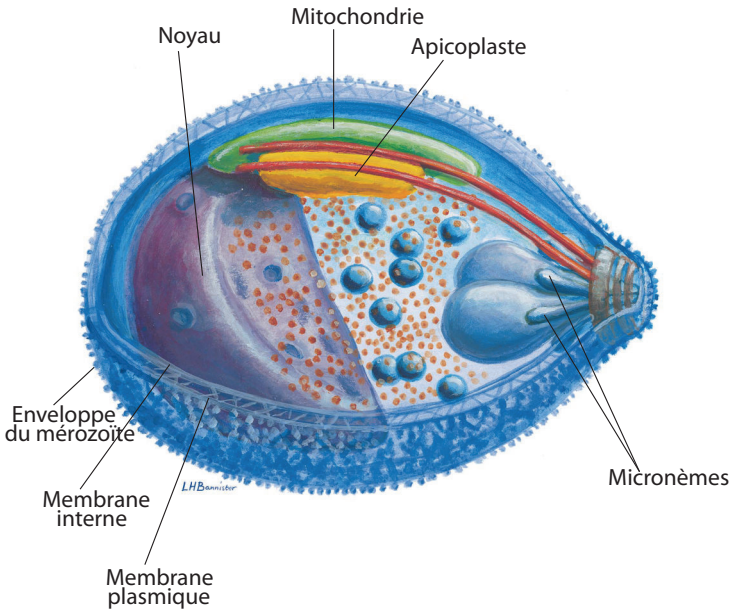



**Figure 4 : Morphologie d'un mérozoïte (stade sanguin) de *Plasmodium falciparum* (Bannister, *et al., 2000).*** Quelque soit son stade, le parasite dispose d'un noyau, d'une mitochondrie unique et d'un apicoplaste. Les micronèmes ont pour rôle de libérer les protéines nécessaires à l'invasion des cellules hôtes.

Après avoir traversé plusieurs cellules du foie, le sporozoïte envahit un hépatocyte, s'installe et se multiplie. A ce stade (schizonte hépatique), le parasite se trouve entouré d'une membrane parasitophore. Lorsque celle-ci se rompt, les mérozoïtes qu'elle contient sont libérés et se propagent dans le sang. Le mérozoïte (Figure 4), contient tout ce qui est nécessaire pour infecter les globules rouges. Après l'invasion de l'érythrocyte, il se transforme en trophozoïte. Le trophozoïte va profiter des ressources de l'érythrocyte pour augmenter sa taille, et passer au stade schizonte. Le noyau se divise, les membranes sont synthétisées pour

former entre 8 et 16 nouveaux mérozoïtes, qui seront libérés dans le sang (Bannister, *et al.*, 2000). Cette libération de parasites dans la circulation sanguine provoque les symptômes caractéristiques du paludisme (accès palustre).

Une petite partie des trophozoïtes va se différencier en gamétocytes (mâles et femelles). Cette différenciation est induite par des facteurs environnementaux qui déclenchent des changements épigénétiques régulant la gamétogenèse (Guttery, *et al.*, 2015). Lors du repas sanguin ces gamétocytes sont ingérés par un moustique. Il sont activés par la forte baisse de la température ambiante, l'augmentation du pH, et la présence d'acide xantunéric (Billker, *et al.*, 1997) de ce nouvel hôte. La fusion des gamétocytes mâles et femelles se produit alors dans l'intestin du moustique (Invergo, *et al.*, 2017). Le zygote ainsi formé se transforme en oocinète, une forme capable de traverser les cellules du moustique pour quitter rapidement l'intestin, et ainsi se protéger du microbiote du moustique (Habtewold, *et al.*, 2015). L'oocinète sera transformé en ookyste, qui se multiplie pour produire des milliers de sporozoïtes, qui à leur tour seront libérés dans l'hémolymphe du moustique (équivalent au système sanguin des vertébrés). Les sporozoïtes atteignent alors les glandes salivaires, ou ils sont stockés avant de recommencer le cycle.

## 3. Les différents génomes de *Plasmodium*

Comme dans la forme mérozoïte (Figure 4), tous les stades du parasite contiennent un noyau, une seule mitochondrie, et une organelle appelée apicoplaste, et sont enveloppés dans trois membranes (interne, externe et enveloppe ou encore appelée membrane plasmique).

### *3.1. Génome nucléaire*

Le génome nucléaire de *P. falciparum* est haploïde pendant tout son cycle, sauf lors de la formation du zygote, dans l'intestin du moustique (Obado, *et al.*, 2016). Il fait 23 Mb et est organisé en 14 chromosomes. Il contient près de 5500 gènes dont la majorité code pour des protéines. Parmi ces protéines, 550 ont une séquence signal à destination de l'apicoplaste, et 246 ont une séquence qui les dirige vers les mitochondries. Cependant, seulement 40 % des protéines plasmodiales ont une fonction prédite (Gardner, *et al.*, 2002, Obado, *et al.*, 2016).

### *3.2. Génome mitochondrial*

Composé uniquement de 6 kb, c'est le génome mitochondrial le plus simple décrit à ce jour. Il code pour seulement 3 protéines impliquées dans la chaîne respiratoire, donc essentielles : les sous-unités I et III de la cytochrome c oxydase (genes *cox1* et *cox3*), et l'apocytochrome b (*cob*) (Figure 5). Ce génome contient également des fragments d'ARN ribosomiques, mais aucun gène d'ARNr entier ou d'ARNt (Rehkopf, *et al.*, 2000, Hikosaka, *et al.*, 2013). Ce génome est d'origine procaryote et est transcrit de manière polycistronique. Le transcrit primaire est clivé, et les fragments sont modifiés et traduits à l'intérieur des mitochondries (Rehkopf, *et al.*, 2000, Jackson, *et al.*, 2011).

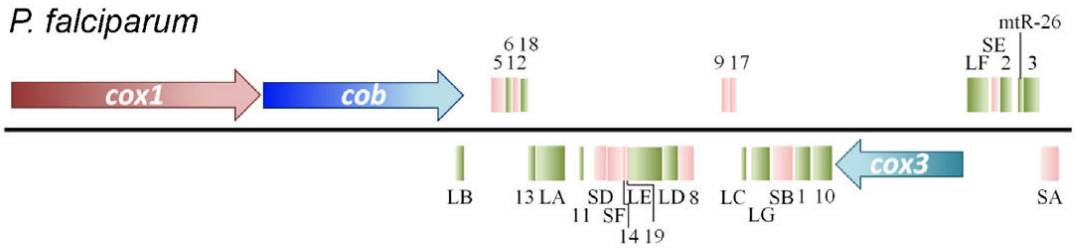



**Figure 5 : Génome mitochondrial de *Plasmodium* (Hikosaka, *et al.*, 2013).** Les flèches épaisses représentent les trois gènes codant pour trois protéines essentielles : les sous-unités 1 et 3 de la cytochrome c oxydase (gènes *cox1* et *cox3*), ainsi que le cytochrome b (gène *cob*). Les zones vertes correspondent à des fragments d'ARNr de la grande sous-unité du ribosome et les zones roses à des fragments d'ARNr de la petite sous-unité du ribosome.

### *3.3. Génome de l'apicoplaste*

Le génome de l'apicoplaste, d'origine procaryote, est constitué d'environ 35 kb d'ADN circulaire (Figure 6) et est très riche en A/T (78 à 86,9 %). Selon les espèces de *Plasmodium*, le génome contient entre 59 et 70 gènes, associés principalement avec sa propre traduction : les ARN ribosomiques (petite et grande sous-unités), des protéines ribosomiques, des protéines chaperonnes et le facteur d'élongation EF-Tu. Il possède également un jeu complet d'ARNt, qui varie entre 27 et 34 gènes selon les espèces (Milton et Nelson, 2016, Arisue et Hashimoto, 2015). Ce génome est transcrit de manière polycistronique. Le transcrit primaire est clivé pour générer les différents ARN (messagers, ribosomiques et de transfert), qui sont utilisés sur place pour la traduction. L'ARNt$^{Leu}$ est le seul transcrit ayant un intron ; il est auto-clivé, comme dans le cas de l'ARNt$^{Leu}$ chez les algues rouges (Nisbet et McKenzie, 2016).

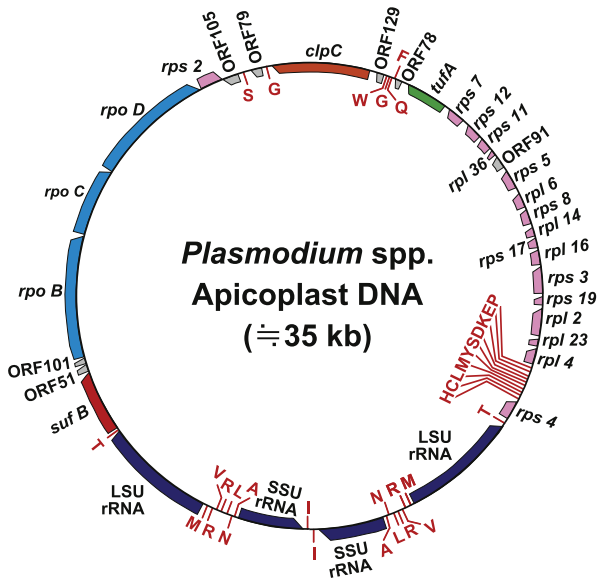



**Figure 6 : Génome apicoplastique de *Plasmodium* (Arisue et Hashimoto, 2015).** Les gènes à l'extérieur du cercle sont transcrits dans le sens horaire, les gènes à l'intérieur dans le sens anti-horaire. Globalement, la même structure est retrouvée chez toutes les espèces, sauf la région dupliquée en bleu foncé, qui ne comporte qu'une seule répétition chez *P. chabaudi*. rps : smal subunit ribosoal protein ; rpl : large subunit ribosomal proteins ; LSUrRNA : Large subunit ribosomal RNA ; SSU : smal subunit ribosomal RNA. Les ARNt sont indiqués par l'acide aminé correspondant.

## 4. L'apicoplaste

L'apicoplaste est une organelle essentielle. Il s'agit d'un plaste qui a perdu sa capacité photosynthétique et qui est présent dans la plupart des *Apicomplexa*. Cependant, il conserve des fonctions typiques des plastes, parmi lesquelles figure la synthèse des acides gras, de l'hème et des précurseurs d'isoprénoïdes. McFadden et Yeh ont étudié la pertinence de ces processus pour le parasite, et ont conclu que la synthèse des précurseurs d'isoprénoïdes est la fonction qui rend l'apicoplaste essentielle pour le parasite (McFadden et Yeh, 2017). Cette voie métabolique est la seule conservée chez tous les parasites qui ont un apicoplaste.

Ce plaste est enveloppé de quatre membranes et a pour origine l'endosymbiose secondaire d'une algue rouge, qui possédait une organelle elle même acquise par endosymbiose d'une algue verte (Figure 7) (Janouskovec, *et al.*, 2010, et revu dans Arisue et Hashimoto, 2015, Milton et Nelson, 2016, Nisbet et McKenzie, 2016, McFadden et Yeh, 2017).

Ce phénomène s'est produit avant la divergence des *Apicomplexa* et des *Dinoflagelata*, il y environ 400 millions d'années (Foth et McFadden, 2003).

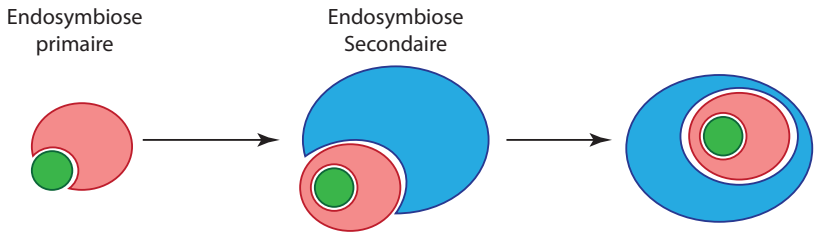



**Figure 7 : Origine évolutive de l'apicoplaste.** En vert, un organisme ancestral de la famille des algues vertes. En rouge, un organisme de la famille des algues rouges. En bleu, l'ancêtre commun aux familles *Apicomplexa* et *Dinoflagellata*.

La conséquence directe de l'origine évolutive de apicoplaste est que ses machineries de transcription et de traduction sont de type procaryote, ce qui rend cette organelle intéressante comme cible pour le développement de nouveaux antipaludéens (Nisbet et McKenzie, 2016). Au cours de l'évolution, cette organelle a perdu la plus grande partie de son information génétique, passant d'environ 150 kb à seulement 35 kb. Une partie des gènes a été transférée au génome nucléaire, comme le gène *Prex*, qui code pour le complexe enzymatique responsable de la réplication et de la réparation de l'ADN apicoplastique (Seow, *et al.*, 2005). Ainsi, plus de 500 des protéines de l'apicoplaste sont traduites dans le réticulum endoplasmique du parasite, et ensuite transportées vers l'apicoplaste grâce à un signal bipartite présent à leur extrémité N-terminale (Gardner, *et al*., 2002). Ce double signal est dû à l'origine évolutive du plaste. Les gènes ont été transférés deux fois au génome nucléaire, après la première et après la deuxième endosymbiose (Foth et McFadden, 2003).

## 5. Interaction hôte/parasite : échanges moléculaires

Au cours du stade sanguin, le mérozoïte envahit un globule rouge et s'installe dans une vacuole parasitophore pour se multiplier. Cette étape, métaboliquement très active, nécessite de nombreux nutriments que le parasite ne synthétise pas. Par conséquent, *Plasmodium* établit une série de voies de transport qui lui permettent d'importer ces nutriments à travers la membrane érythrocytaire, la membrane de la vacuole et sa propre membrane (revu dans

Kirk et Lehane, 2014). Cependant, le fonctionnement exact de ces transporteurs n'est pas encore connu.

### *5.1. Endocytose de l'hémoglobine*

L'Hémoglobine de l'érythrocyte est la source principale d'acides aminés du parasite. L'internalisation de l'hémoglobine est réalisée par endocytose, en utilisant des structures appelées cytostomes (Milani, *et al., 2015*). Ils permettent la formation de vésicules pour internaliser l'hémoglobine. Ils sont localisés au niveau des membranes du parasite et de la vacuole parasitophore et sont associés au système actine/myosine de l'érythrocyte (Figure 8).

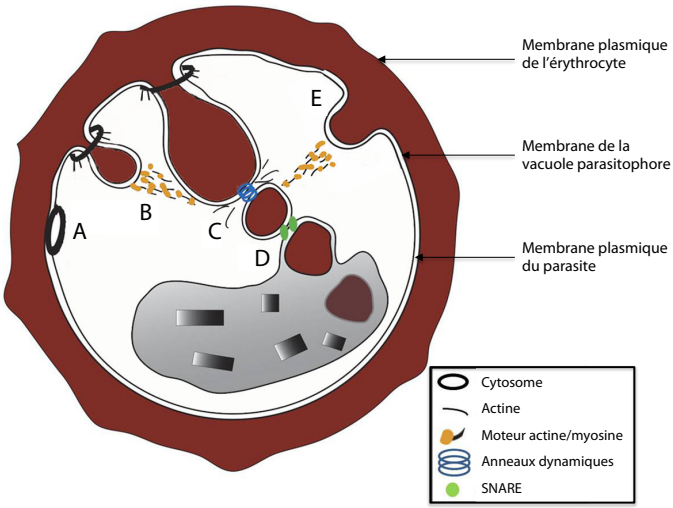



**Figure 8 : Endocytose de l'hémoglobine par un trophozoïte de *Plasmodium* (Milani, *et al.*, 2015).** (**A**) Cytosome localisé à l'interface entre la membrane de la vacuole parasitophore et la membrane plasmique du parasite. (**B**) Naissance et croissance (**C**) d'une vésicule contenant de l'hémoglobine grâce à l'action combinée de l'actine et du moteur actine/myosine de l'érythrocyte. (**D**) Fusion de la vésicule avec la vacuole digestive, impliquant les protéines SNARE. (**E**) Internalisation indépendante des cytosomes par la formation d'un phagotrophe.

Les vésicules ainsi formées sont composées de deux membranes lipidiques : celle du parasite et celle de la vacuole parasitophore. Dans un deuxième temps, les protéines de la famille SNARE (Soluble N-éthylmaleimide-sensitive-factor Attachment protein REceptor) s'assemblent en complexes trans-SNARE pour fusionner ces vésicules à la vacuole digestive du parasite où l'hémoglobine est dégradée et utilisée comme source d'acides aminés (Revu dans Bombardier et Munson, 2015). Il existe un mécanisme d'endocytose de l'hémoglobine, indépendant des cytostomes. Il s'agit de compartiments phagotrophes, c'est à dire de grandes

vésicules indépendantes de l'actine qui se forment tardivement dans les trophozoïtes et les schizonts bien avancés dans leur maturation (Elliot, *et al.*, 2008). En 2015, Milani et coll., favorisait un modèle dans lequel le transport de l'hémoglobine vers le la vésicule digestive se fait principalement par les cytostomes, avec une contribution secondaire possible des phagotrophes, étant donné la faible fréquence des phagotrophes (environ 0,25 par globule rouge infecté) comparée à celle des cytostomes (1 à 4 par globule rouge infecté) au stade trophozoïte, où la majorité de la dégradation de l'hémoglobine se produit (Milani, *et al*., 2015).

### *5.2. Mise en place de la machinerie d'import des nutriments dans le globule rouge infecté*

L'hémoglobine n'est pas la seule molécule nécessaire à la survie du parasite. Pour obtenir des nutriments essentiels, tels que le fer, les purines, les nucléosides, les hexoses et les carboxylates, le parasite utilise soit (1) les transporteurs de l'érythrocyte, comme GLUT1 et hENT1 pour importer le glucose et les purines, respectivement et (2) des protéines synthétisées par le parasite et exportées dans les membranes de la vacuole ou du globule rouge (Milani, *et al.*, 2015, et revu dans Landfear, 2011 et Kirk et Lehane, 2014). En effet, Lorsque le parasite infecte un globule rouge, il remodèle complètement sa membrane et celle de la vacuole parasitophore en y insérant entre autres des transporteurs à large spectre.

Ces transporteurs sont synthétisés dans le réticulum endoplasmique du parasite et sont caractérisés par la présence d'une séquence appelée "*Plasmodium* EXport ELement", ou PEXEL. Ces protéines s'insèrent soit directement dans la membrane de la vésicule parasitophore, soit traverse cette première barrière pour s 'insérer dans la membrane plasmique du globule rouge. Pour traverser la membrane de la vacuole parasitophore, les protéines du parasite utilisent le translocon, appelé aussi PTEX (Koning-Ward, *et al*., 2009) (Figure 9). PTEX est composé de plusieurs sous-unités : (i) EXP2 est la seule protéine membranaire et est responsable de la formation du pore transmembranaire ; (ii) HSP101 est une ATPase dont la fonction est de fournir l'énergie pour le transport ; (iii) PTEX88 et (iv) PTEX150 sont des protéines spécifiques à *Plasmodium* dont les fonctions sont encore inconnues ; (v) TRX2 est une thiorédoxine qui pourrait être impliquée soit dans le dépliement de la protéine avant de traverser la membrane soit dans la stabilisation du complexe PTEX (Revu dans Spillman, *et al*., 2015). Cependant, certaines protéines sont exportées en absence

de la séquence PEXEL, suggérant l'éxistance d'une voie d'export supplémentaire non identifiée (Revu dans Spielman and Gilberger, 2009).

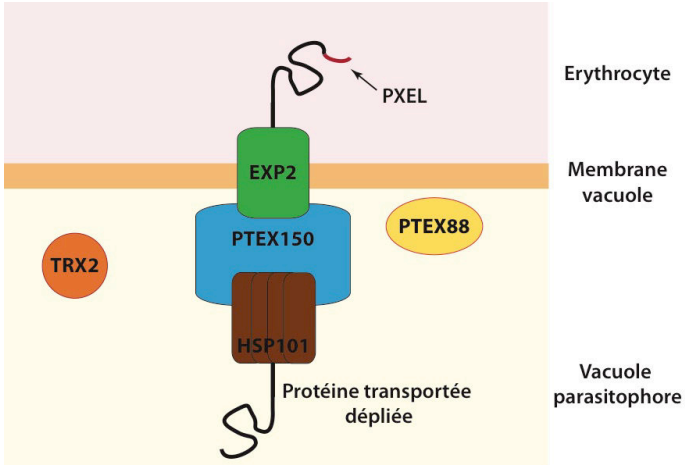



**Figure 9 : Représentation du complexe PTEX de *Plasmodium.*** Figure inspirée de Spillman, *et al.*, 2015.

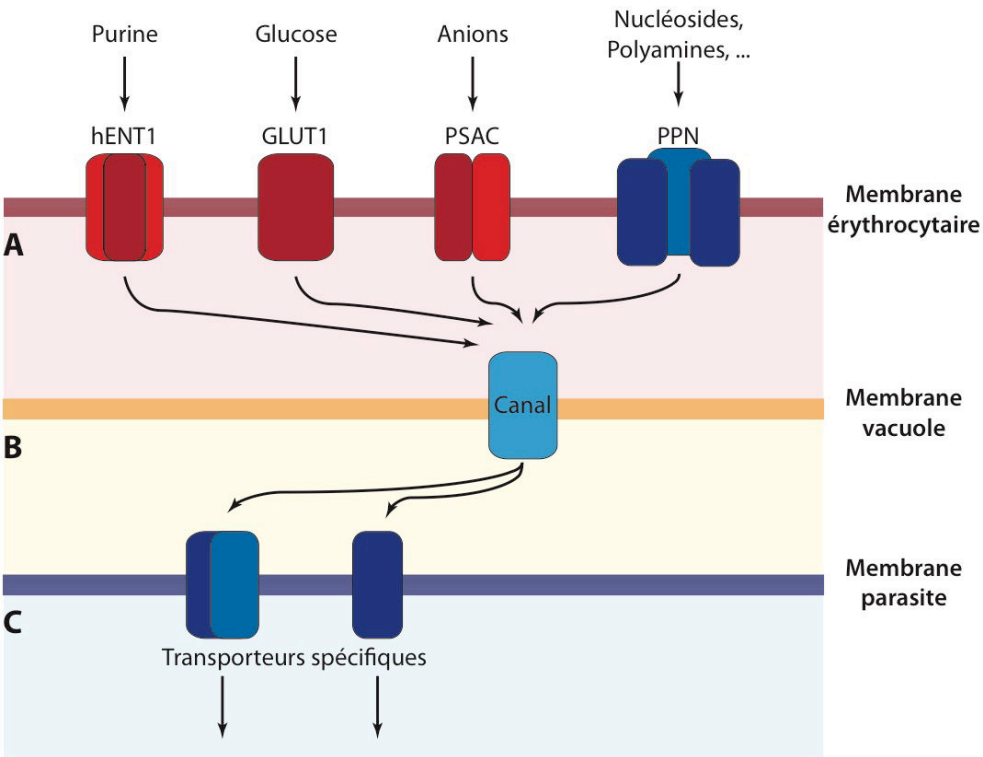



**Figure 10 : Machinerie d'échanges moléculaires entre *Plasmodium* et l'érythrocyte.** Les protéines d'origine plasmodiales sont en bleu, et celles d'origine érythrocitaire sont en rouge. Les transporteurs présents à la surface **(A)** de l'érythocyte, **(B)** de la vacuole parasitophore et **(C)** du parasite permettent à *Plasmodium* d'importer tous les métabolites essentiels à son développement dans le globule rouge.

Grâce à PTEX, plusieurs transporteurs vont être synthétisés dans le parasite et localisés dans la membrane plasmique du globule rouge. C'est le cas de PSAC (*Plasmodium* surface anion channel) et de PPN ("New Permeation Pathway") (Figure 10A). Leur présence augmente la perméabilité de la membrane érythrocytaire rapidement après l'invasion de *Plasmodium* (Ginsburg, *et al*., 1985). Il a été suggéré que ces protéines forment des pores d'environ 0,7 nm dans la membrane érythrocytaire, grâce auxquels, les molécules hydrophobes de petite taille et les anions traversent la membrane du globule rouge. En revanche les cations sont exclus et ne pénètrent pas dans le cytoplasme du globule rouge. Cette spécificité de charge des transporteurs vise à maintenir l'équilibre osmotique à l'intérieur de l'érythrocyte (Ginsburg, *et al*., 1985).

Dans un deuxième temps, les nutriments doivent passer à travers la vacuole parasitophore. Ils utilisent alors un canal d'origine parasitaire qui est fortement représenté dans la membrane de la vacuole (Figure 10B). Ce canal fonctionne comme un tamis moléculaire et ne laisse passer que les molécules dont la taille est inférieure à 1400 Dalton, ce qui concerne la majorité des nutriments du plasma mais aussi des précurseurs métaboliques comme l'ATP synthétisés par la cellule hôte.

Enfin, ces solutés traversent la membrane plasmatique du parasite *via* des transporteurs spécifiques (Figure 10C) (Desai, *et al*., 1993).

## II. LES ARN DE TRANSFERT

A titre d'exemple, le génome humain possède plus de 500 gènes codant pour les différents ARN de transfert (ARNt) (Iben et Maraia, 2012). Il s'agit de l'acide nucléique non codant majoritaire dans les cellules. Il représente entre 4 et 10 % de l'ARN total. Les ARNt sont présents dans tous les domaines de la vie et sont essentiels à la survie des cellules puisqu'ils transportent les acides aminés au niveau du ribosome et permettent la traduction des ARN messagers (ARNm) en protéines. Les aminoacyl-ARNt synthétases (aaRS) catalysent l'estérification des acides aminés sur l'extrémité 3' des ARNt lors d'une réaction très spécifique, l'aminoacylation. Les ARNt interagissent avec une multitude de protéines lors de la traduction et sont également impliqués dans d'autres processus cellulaires.

### 1. Structure des ARNt

La longueur des ARNt cytoplasmiques varie entre 73 et 96 nucléotides, répartis dans cinq domaines, qui forment la structure secondaire dite en « feuille de trèfle » (Figure 11A). Le bras accepteur possède un nucléotide non apparié, appelé base discriminatrice (position 73), et une séquence CCA, constante, il est généralement constitué de 7 paires de bases (pb) Les seules exceptions étant l'ARNt$^{His}$ (Orellana, *et al.*, 1986) et l'ARNt$^{Sec}$ (Sturchler, *et al*., 1993) qui possèdent 8 pb. C'est sur l'adénosine 3' terminale que sera fixé l'acide aminé lors de la réaction d'aminoacylation. Les domaines T et anticodon sont tous deux constitués d'un bras de 5 pb et d'une boucle de 7 nucléotides. Le domaine D quant à lui présente des variations, puisque le bras contient 3 ou 4 pb, et la boucle présente entre 7 et 11 résidus. Finalement, la région variable correspond comme son nom l'indique à une région dont la taille varie de 4 à 21 nucléotides. Cette variation a permis de classer les ARNt en deux familles. Dans la majorité des cas cette région comprend 4 ou 5 résidus ; ce sont les ARNt de la classe 1. Quand la région variable possède entre 13 et 21 résidus (cas des ARNt$^{Tyr}$ procaryotiques, $^{Ser}$, $^{Leu}$, et $^{Sec}$) ils font partie de la classe 2 (Rich et RajBhandary, 1976, Giegé *et al*., 2011).

Afin d'assurer le repliement 3D homogène des ARNt, certains résidus sont conservés et semi-conservés (Figure 11B). En effet, grâce à des interactions à longue distance entre ces

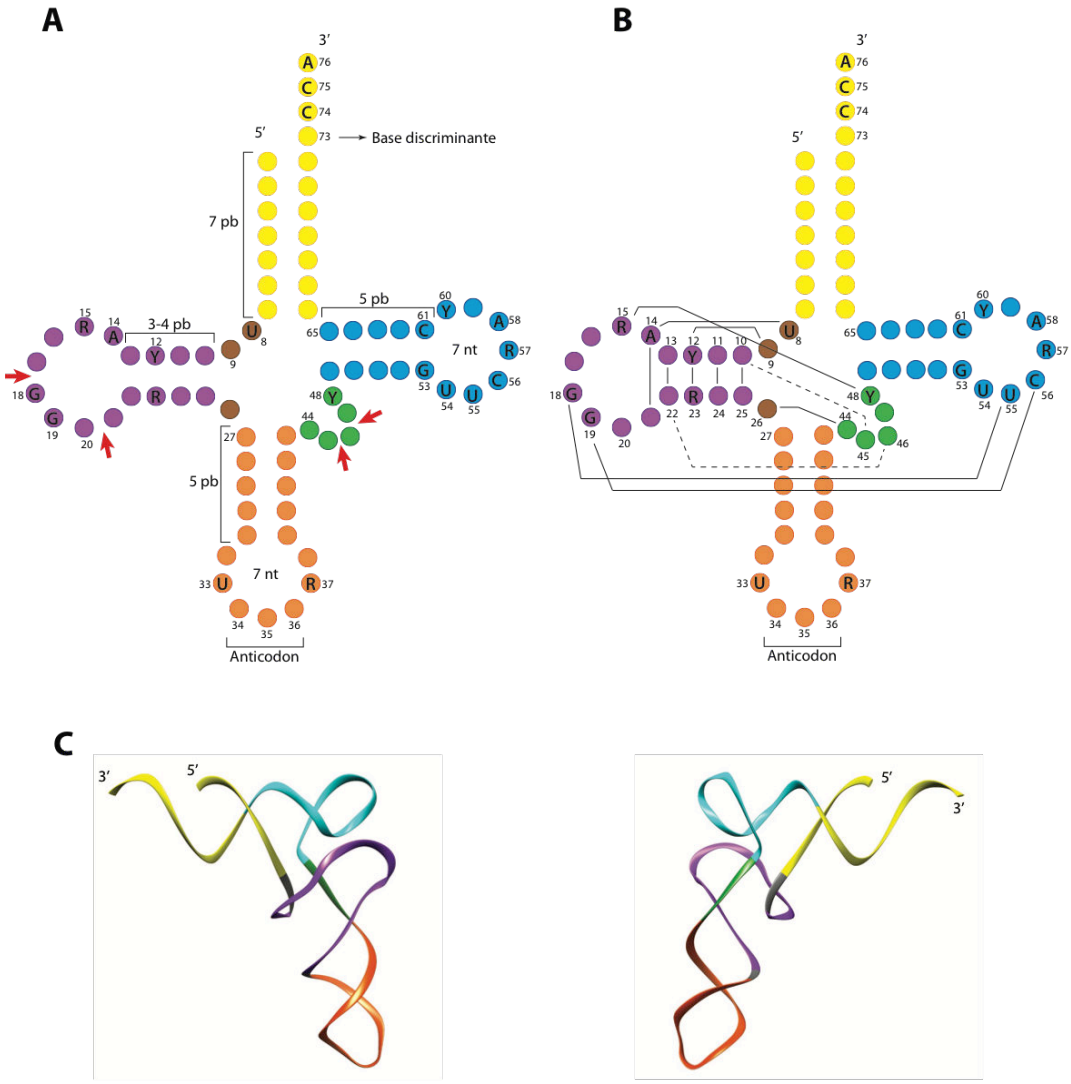



**Figure 11 : Structures secondaire et tertiaire des l'ARNt. (A)** Repliement classique en feuille de trèfle. Les variations de taille sont indiquées par les flèches rouges. **(B)** Représentation des interactions dans la structure secondaire qui permettent **(C)** le repliement en forme de L caractéristique des ARNt. Le bras accepteur est représenté en jaune, le bras D en mauve, le domaine de l'anticodon en orange, la région variable en vert, et le domaine T en bleu. Le nombre de nucléotides dans les boucles et les bras sont indiqués, ainsi que les positions des résidus conservés ou semi conservés. R (purine) et Y (pyrimidine).

résidus, les ARNt se replient tous sur la forme d'un L caractéristique (Figures 11C). Ces interactions tertiaires font intervenir les résidus du bras D ($U_8$, $N_9$, $N_{10}$, $N_{11}$, $Y_{12}$, $N_{13}$, $N_{22}$, $R_{23}$, $N_{24}$, $N_{25}$, $N_{26}$), des nucléotides dans la boucle D ($R_{15}$, $A_{14}$, $G_{18}$, $G_{19}$, $A_{21}$), dans la région variable ($N_{44}$, $N_{45}$, $R_{46}$, $Y_{48}$ dans les ARNt appartenant au groupe 1, et $N_{44}$, $Y_{48}$ dans les ARNt du groupe 2), et dans la boucle T ($\Psi_{55}$, $C_{56}$ et $R_{57}$) (Figure 11B). Certains de ces nucléotides, comme le résidu $\Psi_{55}$, sont modifiés post-transcriptionnellement de manière systématique, comme expliqué dans le paragraphe 3. Cette structure tridimentionnelle en forme de L (Figure 11C)

est strictement conservée, bien que certains ARNt mitochondriaux (mt-ARNt) ne se replient pas en feuille de trèfle classique (Giegé, *et al.,* 2011, Jackman et Alfonzo, 2013). Ceci a été montré pour l'ARNt$^{Cys}$ mitochondrial de l'Armadillo et l'ARNt$^{Tyr}$ mitochondrial de *C. elegans*, qui n'ont pas de bras D et de bras T respectivement (Arnason, *et al*., 1997, Jühling, *et al*., 2010,). Cependant, ils adoptent une structure tridimensionnelle en L similaire grâce à la présence de modifications post-transcriptionnelles à certaines positions (Giegé, *et al*., 2011, Jackman et Alfonzo 2013).

## 2. Maturation des ARNt

La transcription de l'ARNt chez les eucaryotes est réalisée dans le nucléole par l'ARN polymérase III avec les facteurs de transcription TFIIIB et TFIIIC, deux complexes de protéines qui reconnaissent les deux promoteurs internes dans les gènes d'ARNt et recrutent la polymérase (Galli, *et al.*, 1981, Graczyk, *et al*., 2018). Ces promoteurs, appelés boîtes A et B, sont formés par les nucléotides 7 à 19 dans la tige boucle D, et les résidus 52 à 62 dans la boucle T, respectivement (Hofstetter, *et al*., 1981). Lors de la transcription, le pré-ARNt est produit avec des séquences supplémentaires d'environ 12 nucléotides en 5' et en 3' et peut contenir des introns (Revu dans Hopper, 2013). Pour obtenir un ARNt fonctionnel, ce transcrit primaire doit subir un processus de maturation, qui implique l'élimination de ces séquences additionnelles et l'ajout du motif CCA en 3' (Figure 12).

La première étape est la suppression de l'extrémité 5'. Cette réaction est réalisée par la ribonucléoprotéine RNAse P. Dans le cytosol des eucaryotes, ce complexe est composé d'un ARN catalytique et de 10 sous unité protéiques. Cependant, une RNAse P exclusivement protéique a été décrite dans la mitochondrie (Holzmann, *et al.*, 2008, revu dans Schelcher, *et al.*, 2016).

La séquence à l'extrémité 3' est d'abord clivée par la RNase Z, une endonucléase spécifique qui coupe le pré-ARNt immédiatement après la base discriminante ($N_{73}$) (revu dans Phizicky et Hopper, 2010), puis reconstruite grâce à la nucléotidyltransférase appelée «CCA-adding enzyme» ou CCAse. La CCAse ajoute le motif CCA à l'extrémité 3' du pré-ARNt. Elle reconnaît à la fois l'ARNt et les nucléotides ATP et CTP. Elle discrimine un ARNt avec des défauts de structure d'un ARNt correctement replié (Weiner, 2004, Xiong et Steitz 2006,

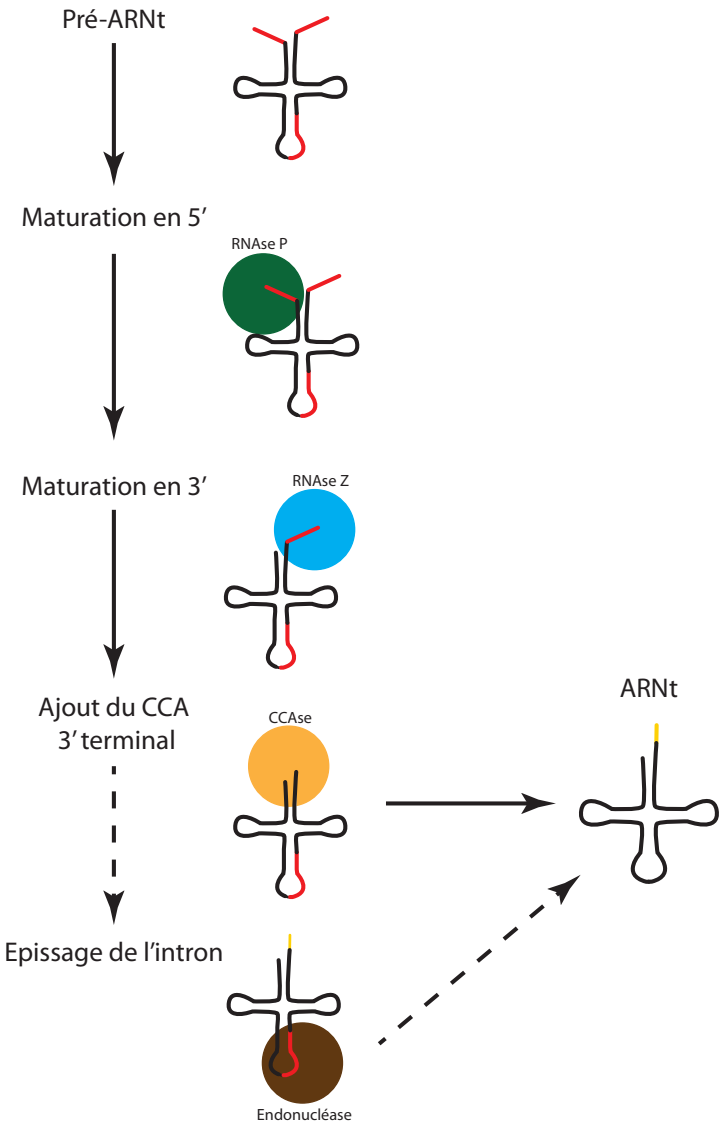



**Figure 12 : Les différentes étapes de maturation des ARNt.** En rouge, les séquences de l'ARNt éliminés pendant la maturation (extrémités 5', 3' et intron), en jaune, le motif CCA 3' terminal.

Betat, *et al.*, 2010, Hou, 2011, Kuhn, 2016). Pendant la réaction, l'ARNt est bloqué dans le site actif de la CCAse. Au fur et à mesure que les nucléotides sont ajoutés, la conformation de l'extrémité 3' à l'intérieur du site actif change pour s'adapter et permettre l'ajout d'un nouveau nucléotide. C'est la position de l'extrémité du pré-ARNt dans le site actif qui détermine le nucléotide à transférer (C ou A) et la fin de la réaction. La présence d'un CCA intègre en 3' est extrêmement importante pour la fonction de l'ARNt dans la traduction, car c'est le site de fixation de l'acide aminé lors de la réaction d'aminoacylation. De plus, comme l'ARNt perdure dans la cellule entre 9 heures et plusieurs jours (Hopper, 2013), le CCA risque

d'être endommagé. La CCase est capable d'identifier les ARNt dégradés et de reconstituer le CCA terminal, elle assure ainsi un rôle de maintien de l'intégrité des ARNt.

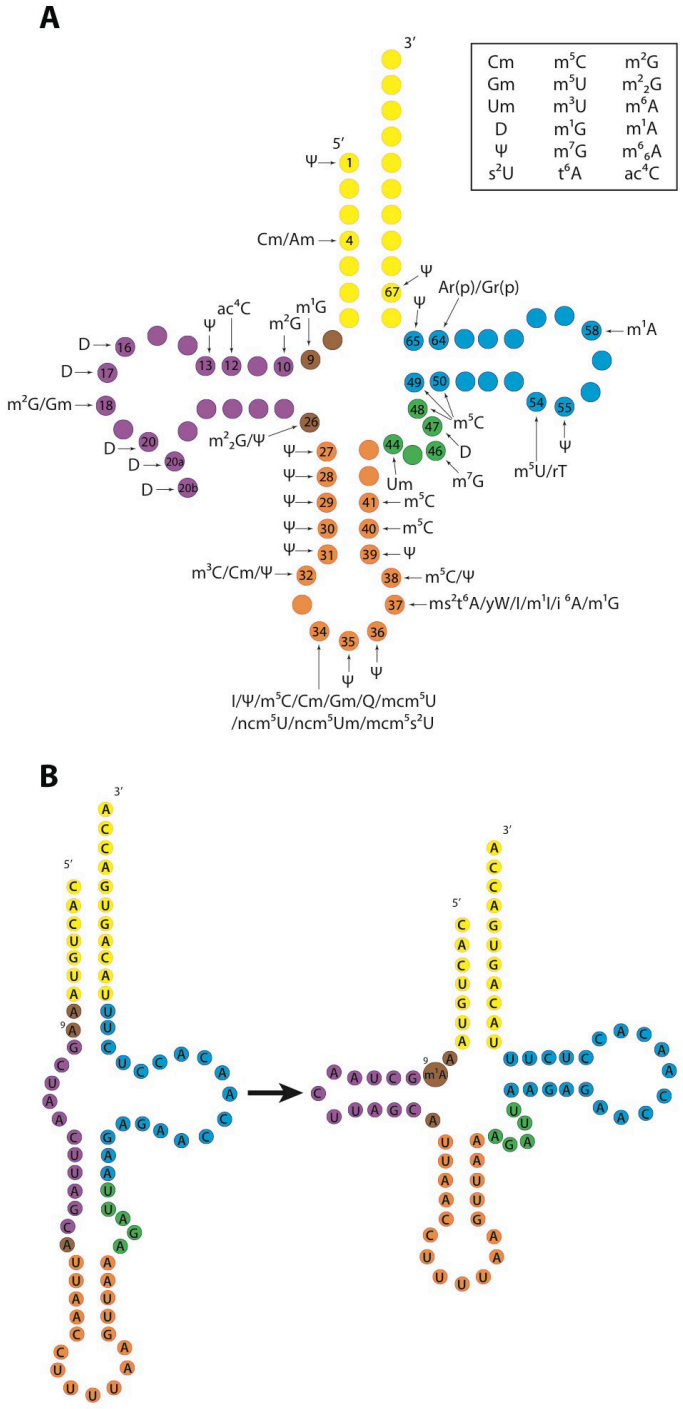
Dans certains cas, comme l'ARNt$^{Phe}$ de levure, le pré-ARNt possède un intron. Située entre l'anticodon et la région variable, cette séquence peut comporter entre 14 et 60 nucléotides. L'épissage se déroule généralement dans le noyau, grâce à une endonucléase non identifiée. Elle reconnaît la structure secondaire en feuille de trèfle, et l'appariement entre le bras de l'anticodon et l'intron (Xue, *et al*., 2006). Cependant, chez la levure, l'épissage est réalisé à la surface de la mitochondrie. Le pré-ARNt doit donc être exporté du noyau vers le cytoplasme pour être dirigé à nouveau dans le noyau après l'épissage (Revu dans Hopper, 2013).

## 3. Modifications post-transcriptionnelles des ARNt

Certains nucléotides sont modifiés post-transcriptionnellement tout le long de la maturation du pré-ARNt. Ces modifications ont été décrites dans le noyau, le cytoplasme, la mitochondrie et les plastes (Revu dans Hopper, 2013, Kirschner et Ignatova, 2015). Les processus de modification peuvent être très complexes. Par exemple, dans le cas particulier de l'ARNt$^{Phe}$ de levure, la modification du nucléotide $G_{37}$ en $yW_{37}$ (wybutosine) est réalisée en plusieurs étapes, par cinq protéines différentes (Tyw1p, Tyw2p, Tyw3p, Tyw4p et Trm5p). La première enzyme à agir est Trm5p. Elle ne méthyle le pré-ARNt dans le noyau qu'après l'épissage de l'intron, or, chez la levure, l'épissage se produit à la surface de la mitochondrie. Le pré-ARNt est alors exporté dans le cytoplasme pour l'épissage, importé dans le noyau, et ré-exporté dans le cytoplasme, ou les enzymes Tyw1p, Tyw2p, Tyw3p et Tyw4p terminent la modification (Ohira et Suzuki, 2011).

Les modifications post-transcriptionnelles consistent à modifier les propriétés physico-chimiques des nucléotides, en ajoutant ou en modifiant des groupes chimiques. Une centaine de modifications différentes ont été décrites dans les ARN, parmi lesquelles, 18 sont conservées dans les ARNt (Figure 13A) (Giegé, 2008, Jackman et Alfonzo, 2013, Kirschner et Ignatova, 2015). Les ARNt sont les acides nucléiques qui accumulent le plus grand nombre des modifications, environ 17 % comparé aux 1 ou 2 % dans l'ARN ribosomal (Jackman et Alfonzo,

2013). En moyenne, un ARNt possède 8 bases modifiées post-transcriptionnellement (Kirschner et Ignatova 2015).




$m^nX$ : méthylation de la base X en position n
$m^n{}_nX$ : diméthylation de la base X en position n
$s^nX$ : l'oxygène en position n dans le nucléotide X est remplacé par un groupe sulfure
Xm : 2'-O méthylation du nucléotide X
D : dihydrouridine
Ψ : pseudouridine
$ac^4C$ : N-4 acetylcytidine
$i^6A$ : N-6 isopentenyladénosine
$t^6A$ : N-6 threonyladénosine
I, inosine (déamination de l'adénosine)
Xr(p) : 2'-O-ribosyl phosphate dérivé du nucléotide X
$mcm^5U$ : C-5 methoxycarbonylmethyl uridine ; $ncm^5U$, 5-carbamoxylmethyluridine ; $ncm^5Um$, 5-carbamoxylmethyl-2'-O-methyluridine ; Q, queosine ; rT, ribothymidine ; yW, wybutosine.

**Figure 13 : Modifications post-transcriptionnelles l'ARNt. (A)** Les modifications post-transcriptionnelles les plus communes sont représentées. L'encadré récapitule les 18 modifications retrouvées au moins une fois dans chaque ARNt. **(B)** Un exemple de la fonction des modifications post-transcriptionnelles : l'effet de la méthylation de l'adénine ($m^1A$) en position 9 sur la structure en feuille de trèfle de l'$ARNt^{Lys}$ mitochondriel humain.

### *3.1. Modifications dans les boucles D et T*

La fonction principale de ces modifications est de stabiliser la structure en forme de L de l'ARNt (Giegé, 2008, Motorin et Helm, 2010, Hopper, 2013). Les modifications les plus abondantes dans les ARNt sont les méthylations ou l'isomérisation de l'uracile, qui forme la pseudouridine (Ψ) (Figure 13A) (Motorin et Helm, 2010, Jackman et Alfonzo, 2013). Certaines sont conservées dans la plupart des ARNt, comme la dihydrouridine (D) aux positions 17 et 20 et la 2'-O-méthylguanosine ($m^2G$) en position 18 dans la boucle D, ou la 5-méthyluridine ($m^5U$) et la Ψ aux positions 54 et 55 respectivement dans la boucle T.

### *3.2. Modifications dans la boucle Anticodon*

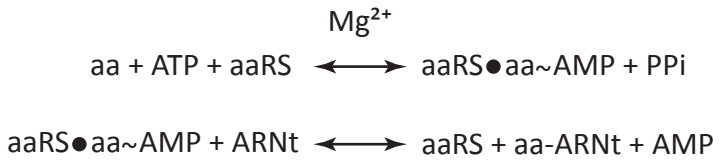
Les positions 34 et 37 dans la boucle de l'anticodon sont très souvent modifiées, mais la nature de cette modification varie considérablement d'un ARNt à l'autre (Figure 13A) (Machnicka, *et al*., 2014). En fait, les modifications de ces deux nucléotides jouent un rôle essentiel dans la traduction. La modification du nucléotide 37 favorise une conformation ouverte et stable de la boucle de l'anticodon, qui facilite l'accessibilité du ribosome à cette région. D'autre part, le nucléotide en position 34 est responsable du wobble, un appariement non-canonique indispensable au décodage de l'ARNm. Une des principales modifications de ce nucléotide est la désamination d'une adénosine, ce qui produit une inosine (I). Ce nucléotide est alors capable de s'hybrider à différents nucléotides (A, C, U). Ainsi, l'interaction efficace du codon avec l'anticodon, et sa traduction, sont assurées (Senger, *et al.*, 1997). Les modifications de la boucle anticodon des ARNt sont donc essentielles pour assurer la fidélité de la traduction. Des mutations dans les gènes des enzymes de modification et/ou dans les gènes des ARNt peuvent entrainer un manque de modification et conduire à des erreurs de traduction graves.

### *3.3. Les modifications dans le reste de la molécule*

Les nucléotides modifiés peuvent diriger le repliement secondaire en forme de feuille de trèfle de l'ARNt en bloquant certains appariements Watson-Crick délétères dans la molécule. C'est le cas de la position 9 dans le l'$ARNt^{Lys}$ mitochondrial humain où la méthylation de l'alanine ($m^1A$) force la formation du bras D (Figure 13B) (Helm, *et al.*, 1999).

## 4. Aminoacylation des ARNt

La fonction principale de l'ARNt est le décodage de l'ARNm pendant la traduction. Cependant, il doit d'abord être aminoacylé, c'est-à-dire qu'un acide aminé doit être chargé sur l'adénosine 3' terminale de l'ARNt. Cette liaison est réalisée par les aminoacyl-ARNt synthétases (aaRS) au cours de la réaction d'aminoacylation, une réaction essentielle et très spécifique. La réaction d'aminoacylation d'un ARNt par son aminoacyl-ARNt synthétase (aaRS) homologue est réalisée en deux étapes :




Au cours de la première étape, un intermédiaire réactif entre l'acide aminé et une molécule d'ATP est formé (Hoagland, *et al.*, 1958). Cette étape, appelée activation, nécessite la présence d'ions $Mg^{2+}$, mais est indépendante de l'ARNt dans la plupart des cas. Cependant, certaines aaRS (ArgRS, GluRS, GlnRS et LysRS de classe I) nécessitent un ARNt pour la formation d'aminoacyl-AMP (aa~AMP). Ensuite, cet acide aminé activé est transféré sur l'OH en 2' ou en 3' de l'adénosine 3' terminale de l'ARNt.

### *4.1. Les aminoacyl-ARNt synthétases*

Les aaRS sont des enzymes modulaires dont la taille et l'oligomérie varient. La présence de certains motifs conservés ont permis de les classer en deux groupes : classe I et de classe II (Tableau 1) (Cusack, *et al.*, 1990, Eriani, *et al.*, 1990, Delarue et Moras, 1993). La seule exception est la LysRS, puisque dans la plupart des bactéries et eucaryotes elle appartiennent à la classe II (LysRS2), alors que le groupe des archaebactéries et certaines bactéries possèdent une LysRS classée dans la classe I (LysRS1) (Ibba, *et al .*, 1997).

La classe I est définie par un site catalytique qui se replie sous la forme d'un Rossman fold contenant deux séquences consensus, KMSKS et HIGH. Les aaRSs de classe I sont essentiellement des protéines monomériques (seulement 3 forment un dimère). Les motifs KMSKS et HIGH interagissent entre eux pour lier et stabiliser l'ATP pendant l'étape d'activation

(Perona, *et al*., 1993). L'autre caractéristique commune à ce groupe de synthétases est la reconnaissance de l'ARNt, et donc l'endroit exact où l'acide aminé est transféré. L'interaction avec l'ARNt se fait du côté du sillon mineur et la fixation de l'acide aminé se fait sur le groupement 2'-OH du ribose de l'adénosine 3' terminale (Delarue et Moras, 1993, Eriani, *et al.*, 1990).

**Tableau 1 : Répartition des aminoacyl-ARNt synthétases en classes et sous-classes**

| | Classe I | | Classe II | | |
|---|---|---|---|---|---|
| Sous-classe | Enzyme | Oligomérie | Enzyme | Oligomérie | Sous-classe |
| Ia | ValRS | $\alpha$ | SerRS | $\alpha_2$ | IIa |
| | IleRS | $\alpha$ | ThrRS | $\alpha_2$ | |
| | LeuRS | $\alpha$ | ProRS | $\alpha 2$ | |
| | MetRS | $\alpha_2$ | GlyRS | $\alpha_2\beta_2$ | |
| | CysRS | $\alpha$ | HisRS | $\alpha_2$ | |
| | ArgRS | $\alpha$ | AlaRS | $\alpha_4$ | |
| Ib | GluRS | $\alpha$ | AspRS | $\alpha_2$ | IIb |
| | GlnRS | $\alpha$ | AsnRS | $\alpha_2$ | |
| | LysRS1 | $\alpha$ | LysRS2 | $\alpha_2$ | |
| Ic | TrpRS | $\alpha$ | PheRS | $\alpha_2\beta_2$ | IIc |
| | TyrRS | $\alpha$ | PylRS | | |
| | | | SepRS | $\alpha_4$ | |

Les synthétases regroupées dans la classe II sont oligomériques. Elles forment surtout des dimères et des tétramères. La seule exception est la PheRS mitochondriale, qui s'organise en monomère (Bullard, *et al.*, 1999). Un motif est conservé dans la zone d'interaction des sous-unités, dans toutes les synthétases appartenant à ce groupe. Ce motif (motif 1) participe à la formation du site d'aminoacylation, ainsi qu'à l'oligomérisation de l'enzyme (Arnez, *et al.*, 1995). De même, deux autres motifs, les motifs 2 et 3, sont également conservés dans ce groupe. Ils participent à la formation du site catalytique (Cusack, *et al*., 1990, Eriani, *et al*., 1990, Ruff, *et al*., 1991). Contrairement aux synthétases de classe I, l'accès au bras accepteur de l'ARNt est effectué par le sillon majeur, de sorte que l'acide aminé est transféré sur le groupement 3'-OH du ribose terminal.

Dans les deux classes, les aaRS sont subdivisés en sous-classes, selon les homologies entre leurs séquences. Il est intéressant de noter que les sous-groupes formés au sein de la classe I reflètent la classification des acides aminés selon leur nature chimique. Ainsi, les enzymes de la sous-classe Ia aminoacylent des acides aminés hydrophobes, celles de la sous-classe Ib, les acides aminés avec des chaînes latérales plus longues, et celles appartenant à la sous-classe Ic, des acides aminés aromatiques.

### *4.2. Spécificité de la réaction d'aminoacylation*

Le décodage correct de l'ARNm dépend non seulement de la reconnaissance entre le codon de l'ARNm et l'anticodon de l'ARNt, mais aussi de la spécificité de la réaction d'aminoacylation. Ainsi chaque aaRS a développé une double spécificité, vis-à-vis de l'acide aminé et vis-à-vis des ARNt homologues. Le site de liaison des acides aminés a une structure rigide qui lie à la fois l'ATP et l'acide aminé spécifiquement dans une position optimale. Ainsi, la spécificité de la reconnaissance est assurée par des interactions qui se limitent à la chaine latérale de l'acide aminé (Revu dans Holler, 1978). De plus, au cours de l'évolution, de nombreuses aaRS ont acquis un domaine supplémentaire (domaine d'editing) qui est capable de corriger l'activation d'un acide aminé erroné par l'enzyme (Revu dans Jakubowski et Goldman, 1992, Giegé et Springer, 2016).

La sélection des ARNt homologues par l'enzyme est gouvernée par la présence d'éléments d'identité. Ces éléments d'identité correspondent à des motifs de séquence et/ou de structure de l'ARNt : soit ils favorisent l'interaction spécifique avec l'aaRS homologue (éléments d'identité positifs ou déterminants), soit ils empêchent l'interaction avec une aaRS hétérologue (éléments d'identité négatifs ou anti-déterminants). Ces éléments sont spécifiques et conservés dans tous les ARNt aminoacylés par une même aaRS : les isoaccepteurs (molécules d'ARNt dont les séquences varient au niveau de l'anticodon mais qui sont porteurs du même acide aminé) comme les isodécodeurs (les ARNt qui partagent le même anticodon, mais dont les séquences varient dans le reste de la molécule).

La description des éléments d'identité est le résultat de nombreuses études. Les revues McClain (1993), Saks et coll. (1994) et Giegé, et coll. (1998) récapitulent les concepts généraux et les données accumulées au fil des ans pour un grand nombre de systèmes bactériens, archaebactériens et eucaryotes. Ces éléments, sont dans la plupart des cas concentrés à

l'extrémité du bras accepteur (la position 73 et les 4 paires de bases précédentes) et dans l'anticodon.

Ils peuvent correspondre à différents motifs, tels que :

- un nucléotide (e.g. la base discriminatrice $G_{73}$ du $tRNA^{Ser}$ ; Soma, *et al*., 1996),

- une paire de bases (e.g. $G_3$:$U_{70}$ de $ARNt^{Ala}$, Hou et Schimmel, 1988),

- un élément structural (e.g. l'orientation de la grande région variable du $tRNA^{Ser}$, Achsel et Gross, 1993),

- une base modifiée (e.g. Inosine en position 34 de l'$ARNt^{Ile}$, Senger, *et al.*, 1997).

Enfin, il est admis que l'identité est globalement conservée au cours de l'évolution pour un même système d'aminoacylation (ARNt/aaRS).

## III. Circulation des ARNt

Après sa synthèse, et tout au long du processus de maturation, l'ARNt fait des allers-retours entre le noyau et le cytoplasme (Figure 14). Malgré sa petite taille, l'ARNt traverse le pore nucléaire par un processus actif (Zasloff, 1983). A la fin du processus de maturation, l'ARNt est localisé dans le cytoplasme pour participer à la traduction des gènes nucléaires. Cependant, selon l'organisme, certains ARNt sont importés dans les mitochondries (Figure 14). En complétant le pool d'ARNt de la mitochondrie, les ARNt importés sont essentiels pour permettre l'expression optimale de son génome. Ces phénomènes ont été présentés dans plusieurs revues (Phizicky et Hopper, 2010, Hopper, 2013, Hopper et Huang, 2015, Chatterjee, *et al*., 2018) dont je fais ici un résumé.

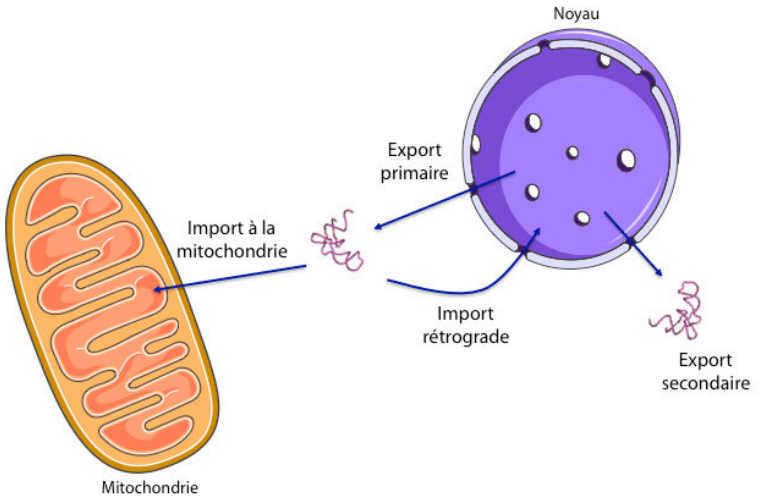



**Figure 14 : Résumé des trajets des ARNt dans la cellule.** Au cours de leur synthèse, les ARNt font plusieurs allers-retours entre le noyau et le cytosol (exports primaire et secondaire, import rétrograde) et certains sont importés dans la mitochondrie pour compléter le pool d'ARNt codés par le génome mitochondrial et participer à la traduction de l'organelle.

### 1. Circulation des ARNt entre le noyau et le cytoplasme

La circulation de l'ARNt du cytoplasme vers le noyau (export primaire, export secondaire ou re-export), et de nouveau vers le cytoplasme (import rétrograde) à travers le pore nucléaire implique différentes protéines nucléaires et cytosoliques (Figure 15).

### *1.1. Transport de l'ARNt du noyau vers le cytoplasme*

Deux protéines responsables du passage de l'ARNt du noyau vers le cytosol ont été décrites chez *S. cerevisiae*, Los1 et Msn5. Elles appartiennent à la famille des $\beta$-importines, et donc, utilisent la voie Ran-GTP pour sortir du noyau avec l'ARNt (Figure 15). La protéine Los1, tout comme ses homologues chez les vertébrés (Export-t ou Xpo-t) (Kutay, *et al.*, 1998) et les plantes (PAUSED) (Hunter, *et al.*, 2003), est impliquée préférentiellement dans l'export primaire des pré-ARNt et des ARNt non aminoacylés dans le cytosol mais est aussi impliqué dans l'export secondaire des ARNt aminoacylés (voir paragraphe suivant) (Hellmuth, *et al.*, 1998). Los1 forme un complexe avec GSP1, une GTPase de la famille Ran-GTP, et l'ARNt pour traverser le pore nucléaire. La présence ou l'absence d'un intron, ainsi que l'aminoacylation de l'ARNt, n'empêchent pas la formation du complexe d'exportation. En effet, il a été montré chez les vertébrés que Xpo-t interagit avec le bras accepteur et les boucles D et T de l'ARNt, mais pas avec le domaine anticodon, zone où se situe l'intron. De plus, bien que les deux extrémités de l'ARNt doivent avoir subi le processus de maturation, cette protéine ne discrimine pas un ARNt aminoacylé d'un ARNt non chargé. Une fois dans le cytosol, la protéine Ran-GTP hydrolyse le GTP en GDP et la protéine Cex1p cytosolique favorise la dissociation de l'ARNt (Nozawa, *et al.*, 2013).

La protéine Msn5 (Takano, *et al.*, 2005) ainsi que son homologue chez les vertébrés (Exportin-5) (Calado, *et al.*, 2002) est impliquée essentiellement dans l'exportation secondaire des ARNt aminoacylés. En exportant uniquement des ARNt matures et aminoacylés, Msn5 participe à une étape de contrôle de qualité essentielle pour la cellule. Ainsi, un ARNt qui n'est pas replié, maturé ou modifié correctement, ne sera pas aminoacylé dans le noyau et ne sera pas exporté dans le cytosol pour participer à la traduction. L'ordre d'assemblage de ce complexe d'exportation secondaire n'a pas été déterminé précisément, mais il a été proposé que Msn5 et GSP1 interagissent avec Tef1/2 (un homologue nucléaire du facteur d'élongation eEF-1$\alpha$ qui stabilise le complexe et permet la liaison à l'ARNt dans la levure) (Huang et Hopper, 2015). Une fois dans le cytosol, eEF-1$\alpha$ prend le relais en liant l'ARNt aminoacylé et en le transportant jusqu'au ribosome (Nozawa, *et al*., 2013).

Alors que l'exportation de l'ARNt du noyau vers le cytoplasme est essentielle pour la survie de la cellule, aucune des protéines présentées ci-dessus ne l'est. Cependant, en absence de Los1, la quantité de l'hétérodimère Mex67-Mtr2 (NXF-NXT ou TAP-p15 chez les

métazoaires) augmente considérablement dans la cellule. Mex67, Mtr2 et leurs homologues ne sont pas des ß-importines, et ce transport est donc indépendant de la voie RanGTP (Chatterjee, *et al.*, 2017).

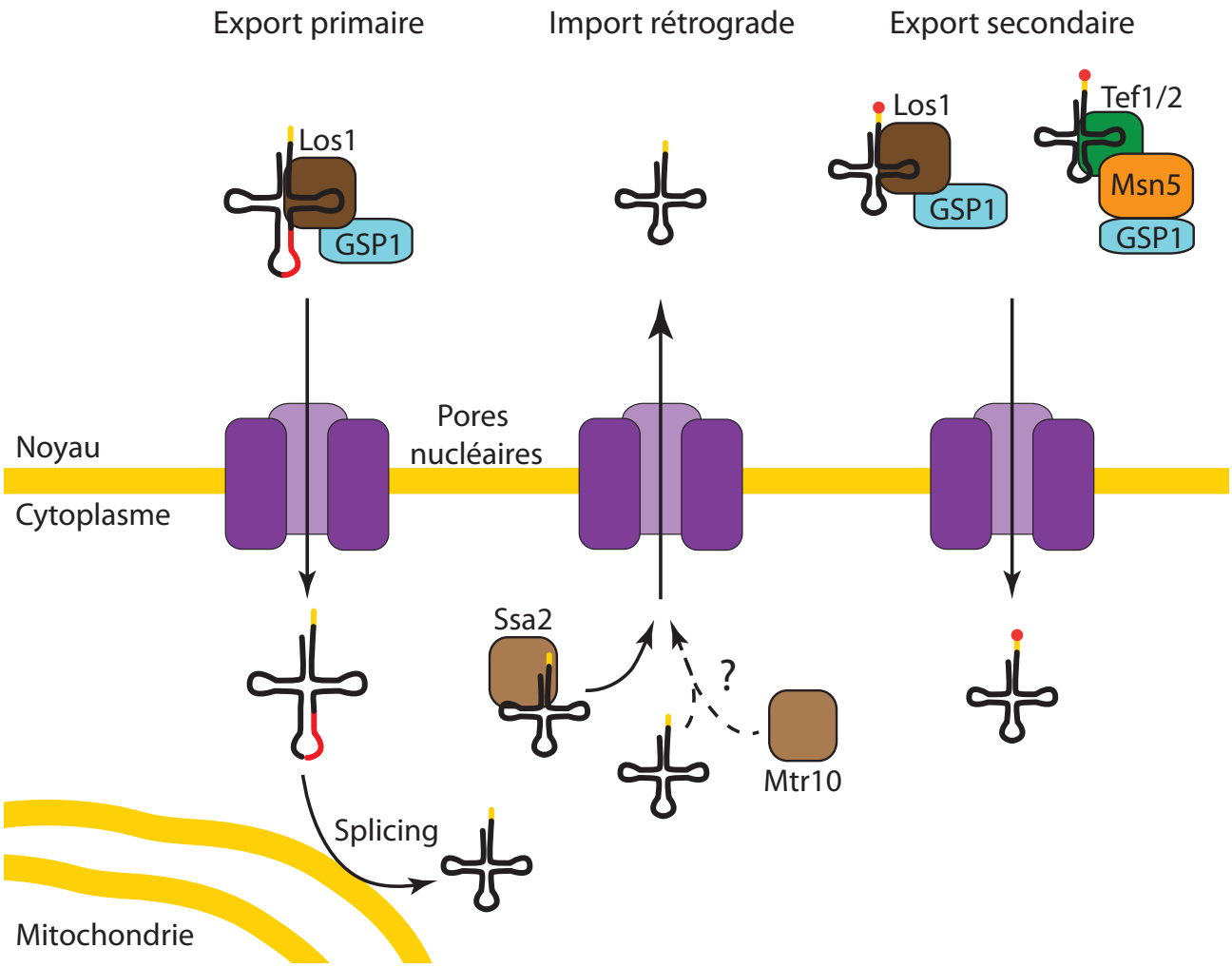



**Figure 15 : Mécanismes de transport de l'ARNt entre le noyau et le cytoplasme chez *S. cerevisiae*.** Lors de l'export primaire, le pré-ARNt possédant un intron (rouge), et la séquence CCA 3' terminal (jaune) traverse le pore nucléaire (mauve). Le mécanisme d'import rétrograde n'a pas encore été décrit, mais les protéines Ssa2 et Mtr10 sont soupçonnées d'y jouer un rôle. Deux voies ont été décrites pour l'export secondaire (Los1 et Mns5). Cet export est essentiel pour convoyer l'ARNt mature et aminoacylé dans le cytosol.

### *1.2. Transport du cytosol vers le noyau*

L'import rétrograde, du cytosol vers le noyau, a été décrit pour la première fois chez la levure, dans des conditions de pénurie de nutriments (Takano, *et al.*, 2005). Dans cette condition de stress, les ARNt matures sont redirigés vers le noyau, pour réguler rapidement la

traduction. Cependant, ce transport rétrograde se produit également de manière constitutive, puisqu'il joue un rôle essentiel au cours de la maturation des pré-ARNt (Ohira et Suzuki, 2011). Cependant ce processus n'est pas bien connu et pourrait impliquer plusieurs mécanismes :

- La protéine Ssa2p qui est une protéine chaperone de la famille Hsp70 qui reconnait directement les ARNt. Les protéines Ssa interagissent avec les composants du pore nucléaire, ce qui facilite leur passage à travers cette structure. De plus, en se liant aux ARNt dont le CCA est partiellement détruit, Ssa2p participe au contrôle qualité (Takano, *et al*., 2015, et revu dans Chatterjee, *et al*., 2018).

- La protéine Mtr10 chez la levure (et son homologue dans les cellules de vertébrés, TNPO3/TRN-SR2) appartient à la famille des ß-importines. Son mécanisme d'action n'a pas encore été décrit et la formation d'un complexe entre Mtr10, l'ARNt et la protéine RanGTP n'a pas non plus été détectée, suggérant que Mtr10 aurait peut être un rôle indirect, tel que l'ancrage des ARNt dans le noyau. Cependant, en absence combinée d'export (Los1 et Msn5) et de Mtr10, les ARNt matures ne s'accumulent pas dans le noyau (Huang et Hopper, 2015,) suggérant que cette protéine serait directement impliquée dans l'import des ARNt.

## 2. Circulation des ARNt du cytoplasme vers la mitochondrie

De la même façon que de nombreuses protéines mitochondriales, certains ARNt codés dans le noyau sont importés dans les mitochondries. En raison de son origine endosymbiotique, cette organelle possède son propre génome qui s'est réduit au fil du temps, que ce soit parce que les gènes ont été transférés vers le noyau, ou parce que les protéines codées ont été remplacées par des protéines nucléaires similaires (revu dans Rubio et Hopper, 2011). L'identification des voies de transport utilisées par les ARNt pour entrer dans la mitochondrie reste un défi. Cette tâche est sans doute compliquée par la grande diversité des mécanismes propres à chaque organisme. À titre d'exemple, certains d'entre eux sont présentés ci-dessous (Figures 16 à 18).

### *2.1. Mécanisme d'import des ARNt dans la mitochondrie de Saccharomyces cerevisiae*

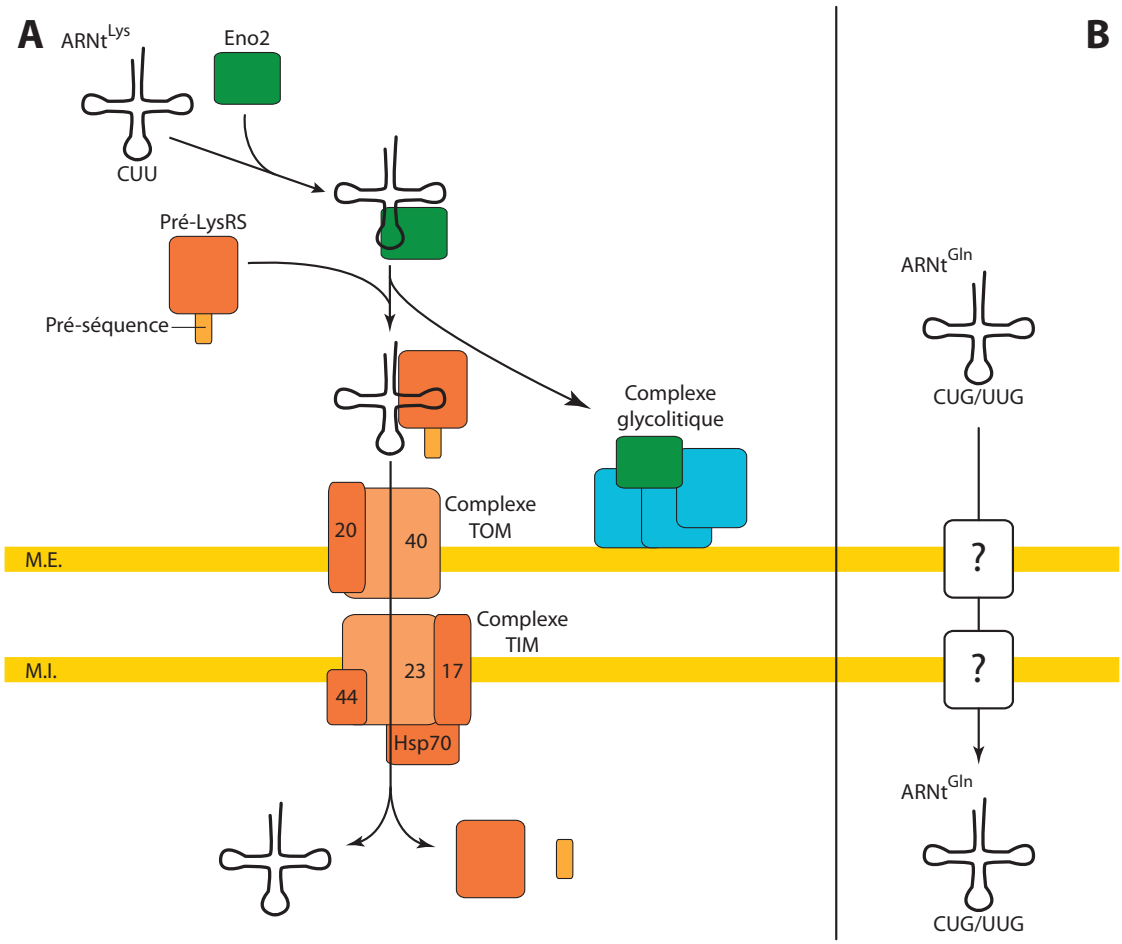



**Figure 16 : Mécanismes d'import des ARNt dans la mitochondrie dépendant (A) et indépendant (B) de cofacteurs cytosoliques dans *S. cerevisiae*.** Les membranes externe (M.E.) et interne (M.I.) de la mitochondrie sont représentés en jaune. Figure adaptée de (Schneider, 2011). Les poids moléculaires des complexes Tom et Tim ont été reproduits de (Salinas-Giegé, *et al.*, 2015).

Un seul système de transport des ARNt cytosoliques a été décrit chez la levure. Ce mécanisme est spécifique de l'ARNt$^{Lys}$(CUU) et implique deux protéines cytosoliques dont les principales fonctions sont sans rapport avec le transport (Figure 16A). Il s'agit d'une énolase glycolytique, Eno2, et du précurseur de la lysyl-ARNt synthétase mitochondriale (mt-LysRS). Dans un premier temps, Eno2 se lie à tRNA$^{Lys}$ et le rapproche de la mitochondrie en se liant au complexe glycolytique localisé sur la membrane externe. Dans un deuxième temps, l'ARNt$^{Lys}$ est pris en charge par le précurseur de la mt-LysRS (traduite localement). Ce complexe utilise alors la voie habituelle d'import des protéines dans les mitochondries, constituée par TOM et

TIM (Translocase Outer/Inner mitochondrial Membrane), complexes localisés respectivement dans les membranes externe et interne (Revu dans Schatz et Dobberstein, 1996). Il a également été montré que l'ARNt$^{Gln}$(CUG) et l'ARNt$^{Gln}$(UUG) sont importés dans les mitochondries (Figure 16B). Dans ce cas, le mécanisme est encore inconnu. On sait seulement que ce transport est indépendant de la voie de translocation des protéines vers les mitochondries, et ne nécessite pas de cofacteurs cytosoliques (Rinehart, *et al.*, 2005).

### *2.2. Mécanisme d'import des ARNt dans la mitochondrie des Kinetoplastides*

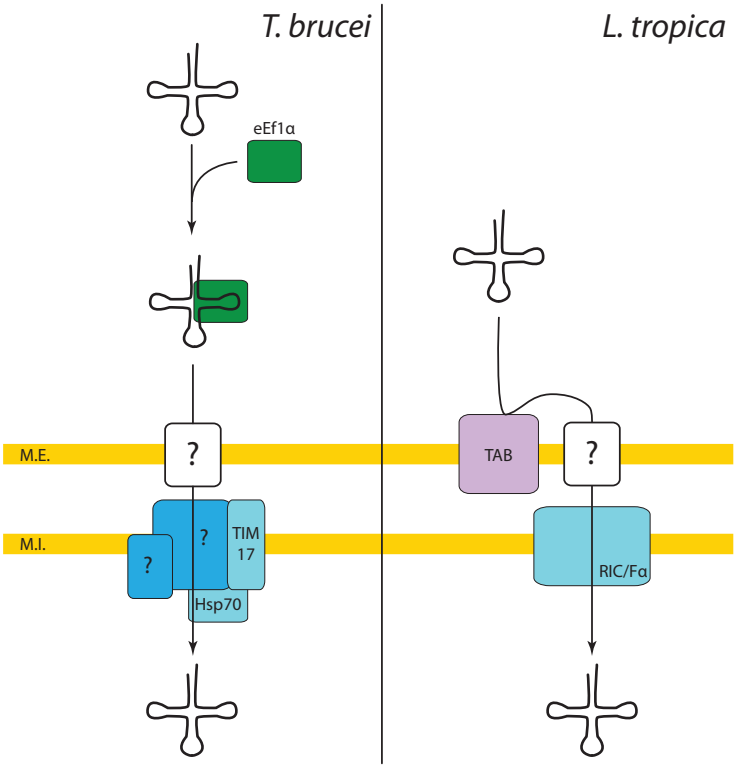



**Figure 17 : Mécanismes d'import des ARNt dans la mitochondrie de *T. brucei* et *L. tropica*.** Les membranes externe (M.E.) et interne (M.I.) de la mitochondrie sont représentés en jaune. Le mécanisme de transport des ARNt chez *T. brucei* à travers la membrane externe n'a pas été décrit. Les deux protéines manquantes (?) du complexe dans la membrane externe sont Tb11.01.4590 et Tb09.v1.0420, mais ni leur fonction ni leur position dans le complexe ont été déterminées. Chez *L. tropica*, l'ARNt interagit avec la protéine TAB, traverse la membrane externe par un mécanisme encore inconnu et ensuite la protéine RIC/Fα permet le passage de l'ARNt à travers la membrane interne.

Dans ces protozoaires, le transport des ARNt dans les mitochondries joue un rôle extrêmement important puisque que leur génome mitochondrial ne contient aucun gène codant pour les ARNt. Par conséquent, la traduction mitochondriale dépend exclusivement de

l'import des ARNt cytosoliques. Les organismes les plus étudiés sont *Trypanosoma brucei* et *Leishmania tropica* (Figure 17). Bien qu'appartenant au même groupe phylogénétique, les voies de transport des ARNt sont différentes dans les deux protozoaires.

Dans le cas de *T. brucei*, le facteur d'élongation eEF-1 $\alpha$ joue un rôle important (Bouzaidi-Taili*, et al.*, 2007). Le facteur eEF-1 $\alpha$ reconnaît la branche acceptrice des ARNt aminoacylés et les dirige vers la membrane externe des mitochondries (Dreher, *et al*., 1999). Le transporteur qui permet de traverser la membrane externe de la mitochondrie reste inconnu, mais des progrès ont été réalisés dans l'identification du mécanisme qui permet la traversée de la membrane interne. La protéine TIM17, impliquée dans le transport des protéines cytosoliques dans les mitochondries, et la protéine chaperone Hsp70 sont directement impliqués dans l'import des ARNt cytosoliques (Tschopp*, et al*., 2010). La présence de Tb11.01.4590 et Tb09.v1.0420 dans le complexe a été décrite (Seidman, *et al*., 2012), mais leur fonction dans le transport des ARNt reste à découvrir.

*Dans L. tropica*, les ARNt cytosoliques ne recrutent pas de cofacteur cytosolique pour leur transport. La première étape de la traversée se fait par un mécanisme encore mal connu. La protéine TAB (Tubular protéine Antisence-Binding), localisée à la surface de la membrane externe des mitochondries, a été proposé comme transporteur, mais son implication doit être vérifiée (Adhya, *et al*., 1997). Cependant, l'ARNt traverse la membrane interne à l'aide du facteur RIC/F $\alpha$ . Cette protéine fait partie du complexe RIC (RNA Import Complex), formé de 11 sous-unités. RIC/F $\alpha$ est impliquée dans la chaîne respiratoire, comme d'autres sous-unités du complexe RIC, tels que la cytochrome c oxydase et la cytochrome c réductase, dont la participation au transport des ARNt est suspectée. RIC/F $\alpha$ reconnaît le motif AGAGY localisé dans le bras D de l'ARNt. Cette reconnaissance spécifique permet à RIC/F $\alpha$ de faire la différence entre les ARNt de classe 1 (région variable de 4 ou 5 nucléotides) et les ARNt de classe 2 (grande region variable) (Goswami, *et al*., 2006).

### *2.3. Mécanisme d'import des ARNt dans la mitochondrie de plantes*

Bien qu'ils aient été mis en évidence dans de nombreuses espèces (*Arabidospis thaliana*, *Nicotiana tabacum*, *Solanum tuberosum*,...), beaucoup de questions sans réponse persistent dans la caractérisation des mécanismes d'import des ARNt dans la mitochondrie chez les plantes (Salinas-Giegé, *et al*., 2015). La seule voie d'import des ARNt décrite à ce jour

nécessite l'énergie apportée par la différence de potentiel de membrane et par l'hydrolyse de l'ATP, et n'implique aucun cofacteur cytosolique (Delage, *et al.*, 2003). L'ARNt utilise d'abord le complexe TOM (Tom20 et Tom40), puis dans un deuxième temps VDAC (Voltage-Dependent Anion Channel) (Figure 18A). VDAC est une porine qui dépend du voltage et qui importe des métabolites tels que l'ATP (Figure 18B). Ce mécanisme reste à étudier en détail, afin d'expliquer comment un ARNt peut traverser un canal prévu pour un nucléotide (Salinas, *et al.*, 2006, Salinas, *et al.*, 2014). Sur la structure de la Figure 18B, on voit qu'une hélice ferme partiellement le pore (en haut à droite). Le diamètre des pores est de 25 Angström aux extrémités mais il pourrait être plus large en déplaçant l'hélice N-terminale hors du canal (en bas à droite).

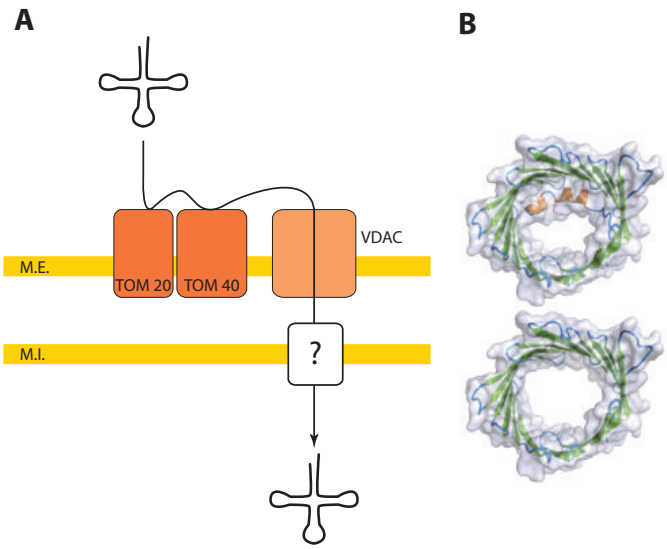



**Figure 18 : Mécanismes d'import des ARNt dans les mitochondries chez les plantes. (A)** Mécanisme général. Les membranes externe (M.E.) et interne (M.I.) de la mitochondrie sont représentés en jaune. L'ARNt interagit avec deux protéines du complexe TOM (TOM20 et TOM40) avant de traverser la membrane externe via le canal VDAC. Le mécanisme de transport à travers la membrane interne n'a pas été décrit. **(B)** Architecture de VDAC. Structure cristallographique de VDAC1 mitochondriale de souris (PDB ID:3EMN). L'architecture de VDAC est caractéristique avec une hélice N-terminale (orange) présente à l'intérieur du canal (Figure adaptée de Salinas, *et al.*, 2014).

L'ensemble de ces observations démontre que les mécanismes d'importation des ARNt connus à ce jour sont très différents. En général, ils sont ATP dépendants, ils impliquent souvent des protéines solubles dont le rôle est de dérouter les ARNt vers la voie d'import. L'import des ARNt détourne des transporteurs qui ne sont pas spécifiques des ARNt ; Il utilise

soit le pore nucléaire, soit VDAC, soit des composants du système d'import des protéines mitochondriales.

## IV. Les ARNt de Plasmodium

### 1. Rapide survol des caractéristiques structurales des ARNt de *Plasmodium*

#### *1.1. Les ARNt de l'apicoplaste*

Les ARNt codés par le génome de l'apicoplaste (api-ARNt), sont riches en nucléotides A et U (Pütz, *et al*., 2010). Les nucléotides conservés et semi-conservés sont retrouvés dans la majorité des api-ARNt, à l'exception des ARNt Trp, Glu, Gly(TCC), Gly(ACC) et Cys (Figure 19). L'ARNt$^{Trp}$ possède un A au lieu d'un G en position 19 dans la boucle D. Dans le cas des ARNt

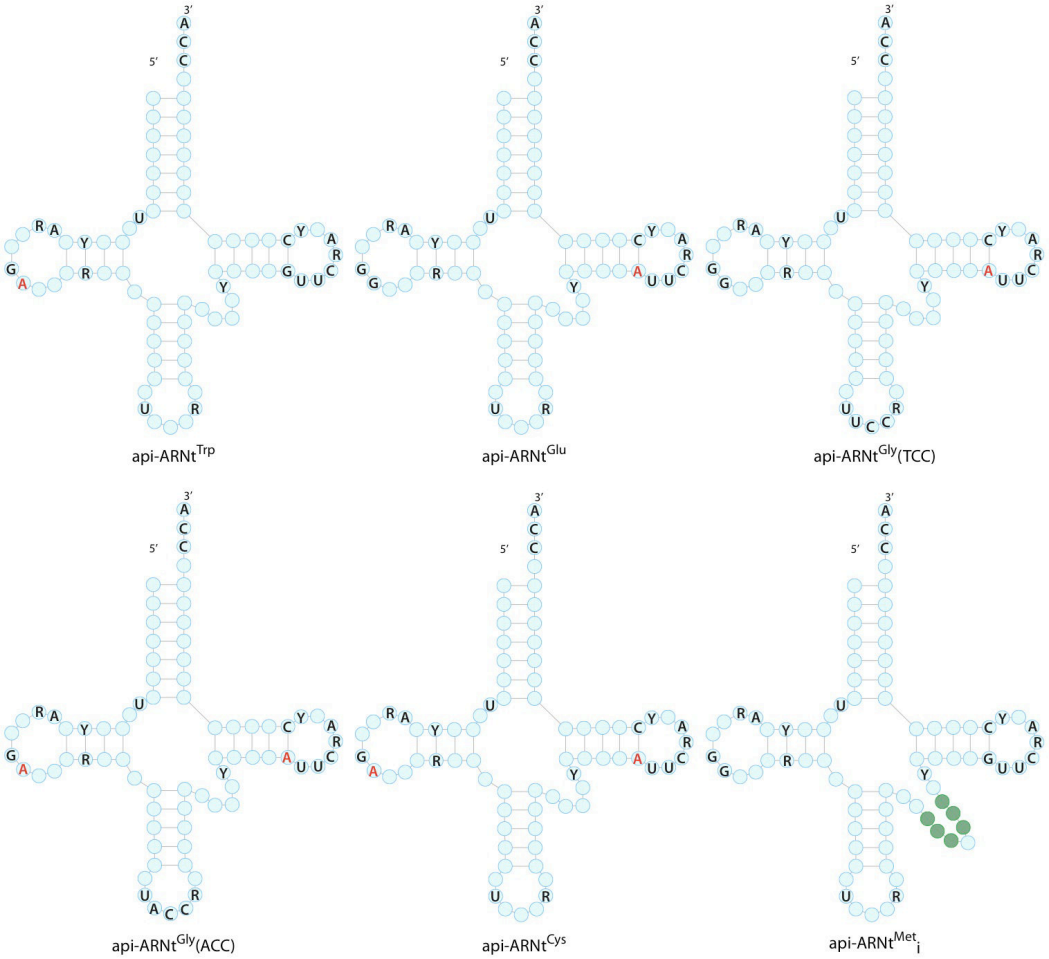



**Figure 19 : Récapitulatif des différences structurales dans les api-ARNt de *Plasmodium*.** En rouge, les nucléotides qui remplacent les résidus conservés ($A_{19}$ et $A_{53}$), et en vert, les nucléotides supplémentaires dans la région variable de l'api-ARNt$^{Met}$ initiateur. Dans ce cas précis on peut également proposer un repliement alternatif de l'ARNt$^{Met}$ initiateur où que les 11 nucléotides supplémentaires font partie d'un intron. Après excision de l'intron, l'ARNt$^{Met}$ initiateur contiendrait une paire de bases G29-G41 dans le bras anticodon et une région variable de 4 nucléotide (ARNt de classe 1).

Glu et Gly(TCC), le motif GUUC dans la boucle T a été transformé en AUUC. Cette mutation a aussi été observée dans des ARNt d'autres plastes. Enfin, les ARNt Gly(ACC) et Cys accumulent les deux mutations ($A_{19}$ et le motif AUUC dans la boucle T). Ces mutations se compensent, ce qui permet de rétablir l'interaction entre les boucles D et T (Preiser, *et al.*, 1995). De plus, une anomalie a été mise en évidence dans l'api-ARNt$^{Met}$ initiateur. Il possède une région variable de 11 nucléotides, alors que les ARNt$^{Met}$ appartiennent à la classe I. Selon les auteurs, cette caractéristique serait propre à ARNt$^{Met}$ initiateur apicoplastique et le distinguerait de l'ARNt$^{Met}$ élongateur (Pütz, *et al*., 2010).

### *1.2. Le cas particulier de la mitochondrie*

Le génome mitochondrial de *Plasmodium* ne code pour aucun gène d'ARNt, comme c'est le cas des génomes mitochondriaux de *Trypanosoma* ou de *Leishmania*. Cependant l'activité de transcription (Rehkopf, *et al*., 2000, Hikosaka, *et al*., 2013) et de traduction (Pino, *et al*., 2010) sont essentielles dans cette organelle. L'import des ARNt cytosoliques dans la mitochondrie de *Plasmodium* a été suggéré pour les ARNt$^{Cys}$ et ARNt$^{Phe}$ et une seule aaRS (PheRS de type prokaryote) (Sharma et Sharma, 2015) mais doit encore être confirmé.

### *1.3. Les ARNt cytosoliques*

Il n'existe pas de différence significative au niveau structural entre les ARNt de *Plasmodium* et ceux des autres eucaryotes. En effet, les ARNt cytosoliques de *Plasmodium* contiennent tous les nucléotides conservés et semi-conservés nécessaires pour leur repliement canonique en forme de feuille de trèfle et en L. En dépit de la forte proportion en A/T du génome de *P. falciparum* (81 %), les ARNt nucléaires présentent un pourcentage de G et de C moyen (environ 57 %), qui correspond à ce qui est observé dans les ARNt humains par exemple.

Etonnamment, le génome de *Plasmodium* ne possède qu'une seule copie de chaque gène d'ARNt, c'est à dire un nombre minimal de 46 gènes (Gardner, *et al.*, 2002 et revu dans Jackson, *et al*., 2011). Ce nombre de gènes est en théorie suffisant pour permettre le décodage de l'ensemble du génome du parasite. Cette particularité est cependant une exception parmi les eucaryotes où on retrouve systématiquement plusieurs copies de chaque

gène d'ARNt, le nombre total de gènes d'ARNt pouvant atteindre 275 copies par exemple chez la levure ou 417 copies chez l'Homme (http://gtrnadb.ucsc.edu/).

### *1.4. Les ARNt et le repliement cotraductionnel des protéines*

La disponibilité des ARNt aminoacylés dans le cytosol des cellules peut influer sur la vitesse de traduction du ribosome. Ce processus favorise le repliement co-traductionnel de la protéine en cours de synthèse (Varenne, *et al.* 1984, Thanaraj et Argos, 1996, Widman, *et al.*, 2008). Les pauses du ribosome permettent le repliement du domaine protéique néo-synthétisé indépendamment du reste de la protéine qui n'est pas encore traduit (Figure 20). En règle générale, il est admis que les codons les plus utilisés correspondent à des ARNt codés par un grand nombre de gènes (Revu dans Novoa et Ribas de Pouplana, 2012). En d'autres mots, dans l'ensemble des organismes, les isoaccepteurs les plus abondants (caractérisés par de multiples copies de gènes) décodent les codons les plus abondants et réciproquement, les isoaccepteurs les plus rares (codés par une seule copie de gène) décodent les codons rares. Ainsi la synthèse de la nouvelle protéine est rapide lorsque le ribosome traduit des codons décodés par les ARNt abondants et elle ralentit lorsque le ribosome traduit une série de codons décodés par les ARNt rares. En observant le génome de *Plasmodium*, il semble que le parasite ne suive pas rigoureusement les mêmes lois.

- *Plasmodium* ne possède qu'un jeu minimal de gènes d'ARNt où chaque isoaccepteur n'est représenté qu'une seule fois. Ainsi, en absence de régulation individuelle de leur transcription, tous les isoaccepteurs seraient représentés de façon identique dans le cytosol.

- L'usage des codons dans *Plasmodium* n'est pas uniforme. Certains ARNt sont extrêmement sollicités. Par exemple un seul gène d'ARNt$^{Asn}$ et un seul gène d'ARNt$^{Lys}{}_{UUU}$ doivent produire suffisamment d'ARNt pour décoder 9,4 % et 8,6% du génome, respectivement (Bastien, *et al.*, 2004). Si les résidus lysine sont répartis équitablement dans les protéines du parasite, ce n'est pas le cas pour l'asparagine. En effet, les résidus asparagine sont regroupés dans des insertions très abondantes dans les protéines de *Plasmodium* (Low Complexity Regions ou LCR) dont on ne connaît pas la fonction.

Ces LCR sont présentent dans 90% des protéines du parasite (Pizzi et Frontali, 2001), et contiennent de longues répétitions d'asparagine. Ainsi, la traduction de ces LCR nécessite de grandes quantités d'Asn-ARNt$^{Asn}$. Il a été proposé que la traduction des LCR riches en

répétitions d'asparagine, ferait de l'ARNt$^{Asn}$ un ARNt rare, réduirait la vitesse de décodage du ribosome localement et pourrait participer au repliement co-traductionnel des protéines de *Plasmodium* (Figure 20) (Frugier, *et al.*, 2010 ; Filisetti, *et al.*, 2013). Cette hypothèse, est la seule qui propose une fonction expliquant l'ubiquité des LCR dans le protéome du parasite.

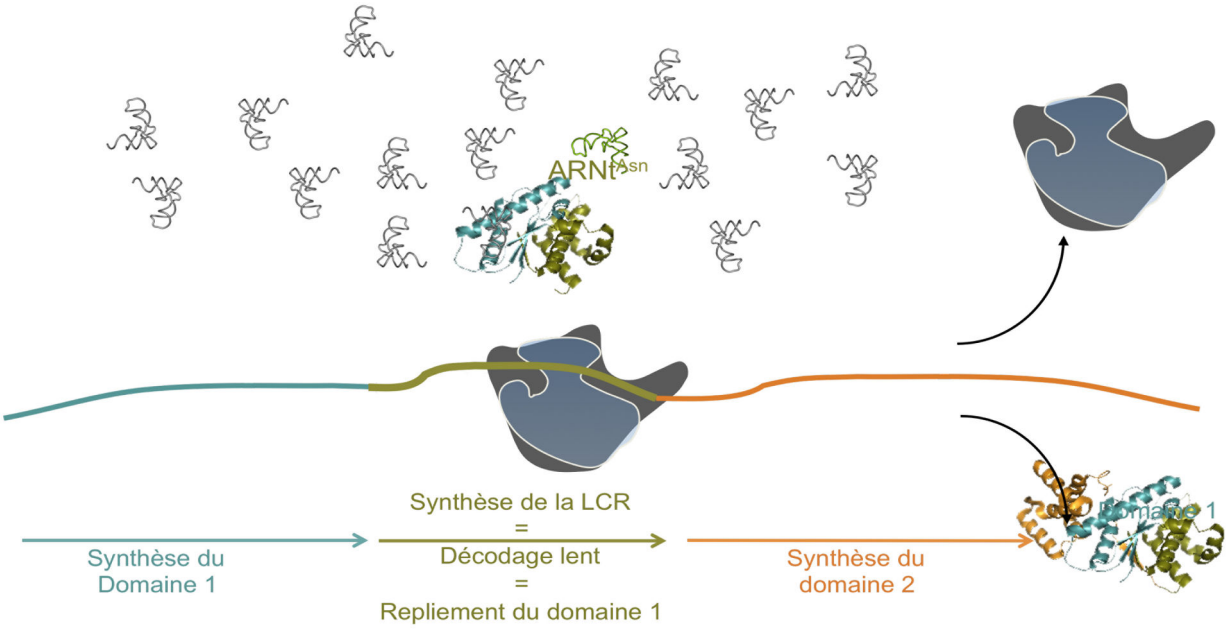



**Figure 20 : Les LCR agissent comme des éponges à ARNt$^{Asn}$.** Résumé du concept décrit dans (Frugier, *et al.*, 2010). Au cours de la traduction d'une protéine constituée de plusieurs domaines structuraux, la concentration limitante en Asn-ARNt$^{Asn}$ dans *Plasmodium* ralentit le ribosome lorsqu'il décode la portion d'ARNm correspondant à une LCR riche en asparagines (en vert). Ce processus donne du temps au premier domaine (en bleu) de se replier indépendamment, avant la synthèse du second domaine (en orange).

Le faible nombre de gènes d'ARNt dans le génome de *Plasmodium* pose également beaucoup de questions quant à la disponibilité des ARNt lors des phases de multiplication intensive du parasite, qui nécessitent une traduction soutenue, au cours de son cycle de vie. Une des possibilités serait que le parasite utilise à moment donné les ARNt de l'hôte (vertébré ou invertébré) pour compenser ce faible nombre de gènes d'ARNt.

## 2. tRip et l'import des ARNt exogène dans *Plasmodium*

Les résultats du laboratoire (Bour, *et al.*, 2016) démontrent l'existence d'un événement unique : les sporozoïtes (forme infectieuse de *Plasmodium* isolée à partir des glandes

salivaires de moustiques infectés) importent des ARNt exogènes *in vitro*. Cet import d'ARNt est rendu possible grâce à une protéine de surface appelée tRip (tRNA import protein) (Figure 21). tRip est composée de 402 acides aminés. Dans son extrémité C-terminale, elle contient un domaine homologue à la protéine Trbp111, un domaine OB-fold considéré comme le module de liaison minimum aux ARNt (Morales, *et al*., 1999). Le domaine N-terminal de tRip est un domaine de type GST et est localisé à l'intérieur du parasite. Une hélice transmembranaire sépare les deux domaines (Figure 21).

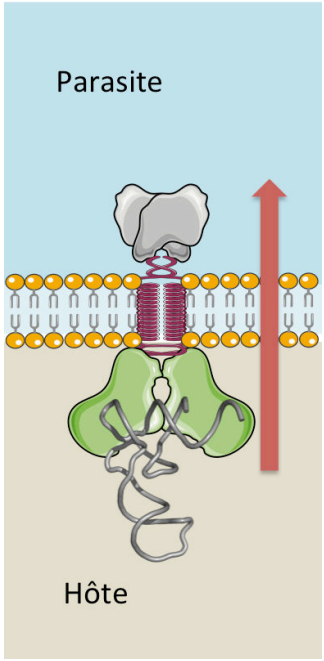



**Figure 21 : tRip et l'import des ARNt exogène dans le sporozïte de *Plasmodium*.** tRip est un homodimère. La protéine est constituée d'un domaine N terminal localisé à l'intérieur du parasite, d'un domaine C-terminal de liaison aux ARNt localisé à l'extérieur du parasite. Les deux domaines sont séparés par une hélice (20 acides aminés) prédite comme étant transmembranaire.

Le gène codant pour tRip n'est trouvé que dans les génomes des *Apicomplexa* et la protéine a été caractérisée : (i) Elle lie spécifiquement les ARNt avec une affinité élevée en reconnaissant leur structure tridimentionnelle ; (ii) tRip est présente à tous les stades du développement du parasite (à la fois dans l'hôte vertébré et dans le moustique) ; (iii) tRip est localisée à la surface du parasite, le domaine liant les ARNt étant exposé à l'extérieur ; (iv) les ARNt exogènes pénètrent rapidement dans les sporozoïtes vivants ; (v) tRip favorise l'entrée de l'ARNt dans le parasite et (vi) le parasite tRip-KO est caractérisé par une synthèse protéique réduite et se développe moins efficacement que le parasite sauvage dans le sang des souris infectées.

Comme l’infection du parasite sauvage est très efficace, les symptômes sont visibles quelques jours après l’injection des parasites dans la souris. En revanche, l’infection par *Pb*tRip-KO est beaucoup plus lente et aléatoire, la quantité de parasites dans le sang peut être très élevée par rapport à un parasite sauvage, suggérant que la virulence de ce parasite mutant est affectée. De plus, une analyse comparative des protéomes des stades sanguins des souches sauvages et tRip-KO permet d'affirmer que l'absence de tRip provoque une altération dans la traduction du parasite muté. La synthèse de certaines protéines importantes pour l'invasion des érythrocytes, entre autres, est diminuée. De manière surprenante, l'expression de certaines protéines augmente en absence de tRip. C'est le cas par exemple de certaines ARNt-aminoacyl synthétases (Thèse de Delphine Kapps, 2016).

## 3. Les domaines OB-fold dans les pathogènes unicellulaires eucaryotes

La protéine tRip est particulièrement intéressante, car c'est la première fois qu'un mécanisme de transport d'ARNt intercellulaire a été décrit. Cette protéine contient un domaine appelé "OB-fold", qui est retrouvé dans de nombreuses protéines et dans tous les organismes. Pour illustrer cette grande variété, les protéines des pathogènes eucaryotes unicellulaires ayant un domaine OB-fold ont été listées, ainsi que leur rôles dans les différents organismes.

FEBS Letters

FEBS PRESS
science publishing by scientists

REVIEW ARTICLE

# OB or Not OB: Idiosyncratic utilization of the tRNA-binding OB-fold domain in unicellular, pathogenic eukaryotes

Delphine Kapps[1], Marta Cela[1], Anne Théobald-Dietrich[1], Tamara Hendrickson[2] and Magali Frugier[1]

1 RNA Architecture and Reactivity, Strasbourg University, CNRS, IBMC, France
2 Department of Chemistry, Wayne State University, Detroit, MI, USA

**Correspondence**
M. Frugier, RNA Architecture and Reactivity, Strasbourg University, CNRS, IBMC, 15 rue René Descartes, 67084 Strasbourg Cedex, France
Fax: +33(0) 388 602 218
Tel: +33(0) 388 417 109
E-mail: M.Frugier@ibmc-cnrs.unistra.fr
and
T. Hendrickson, Department of Chemistry, Wayne State University, 5101 Cass Avenue, Detroit, MI 48202, USA
Fax: +01 313 577 8822
Tel: +01 313 577 6914
E-mail: tamara.hendrickson@chem.wayne.edu

(Received 9 September 2016, revised 28 September 2016, accepted 28 September 2016)

doi:10.1002/1873-3468.12441

Edited by Michael Ibba

**In this review, we examine the so-called OB-fold, a tRNA-binding domain homologous to the bacterial tRNA-binding protein Trbp111. We highlight the ability of OB-fold homologs to bind tRNA species and summarize their distribution in evolution. Nature has capitalized on the advantageous effects acquired when an OB-fold domain binds to tRNA by evolutionarily selecting this domain for fusion to different enzymes. Here, we review our current understanding of how the complexity of OB-fold-containing proteins and enzymes developed to expand their functions, especially in unicellular, pathogenic eukaryotes.**

**Keywords:** parasites; pathogenic eukaryotes; Trbp111; tRNA-binding protein

The roles of most tRNA are to enable protein synthesis. tRNA species have highly conserved L-shaped conformations, wherein the acceptor and T-domains form one arm of the L, while the anticodon and D-domains form the other arm (Fig. 1A). A defining characteristic of the three-dimensional structure of the tRNA is the elbow region, and many enzymes recognize and bind to this region in order to function. In the elbow, the nucleotides of the D- and T-loops interact to give rise to the L-shape of canonical tRNA species [1,2]. The D-loop is composed of 7 to 10 residues that adopt an irregular structure in which the two conserved guanines at positions 18 and 19 form tertiary interactions with residues Ψ55 and C56 in the T-loop (Fig. 1A).

Examples of tRNA elbow recognition include enzymes involved in tRNA maturation, modification, aminoacylation, transport, and translation factors that guide tRNA species and aminoacyl-tRNA species through cell compartments and the ribosome. The diversity of these proteins illustrates various tRNA elbow recognition strategies [3]. Proteins that perform no apparent catalytic function are also known to recognize this elbow region. Notably, an independent and minimal module called Trbp111 (tRNA-binding protein 111) was the first protein characterized in what

**Abbreviations**
GluRS, glutamyl-tRNA synthetase; GST, glutathione S-transferase; LysRS, lysyl-tRNA synthetase.

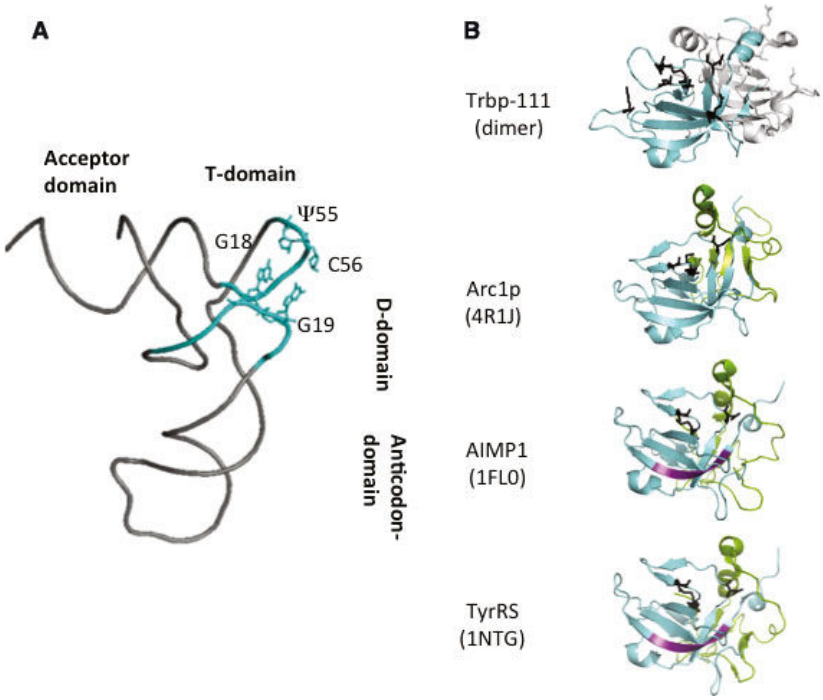



**Fig. 1.** Representative structures of tRNA and tRNA-binding OB-fold domains. (A) Crystal structure of *S. cerevisiae* $tRNA^{Asp}$ [51]. The D- and T-loops are in blue and the four nucleotides G18, G19 (D-loop), pseudouridine Ψ55 and C56 (T-loop) involved in the formation of the tRNA elbow are indicated. (B) Crystal structure of representative OB-fold domains. The *A. aeolicus* dimeric Trbp111 structure [9] is shown with one monomer in blue and the second in gray. The OB-fold domains for the structures corresponding to the *S. cerevisiae* Arc1p [52], human AIMP1 [12], and TyrRS [50] OB-folds are shown. In each case, the Trbp111 conserved domain is in blue and the extended C-terminal domain is indicated in green. The migration motifs (RVGKIIT in AIMP1 and RIGCIIT in TyrRS) are shown in purple. Amino acids identified as essential for tRNA binding (K33, E45, N71, R75, S82, and M85) are indicated in black sticks. The structures were prepared with PyMol (The PyMOL Molecular Graphics System, version 1.8 Schrödinger, LLC. https://www.pymol.org/).

was, at the time, a new class of tRNA elbow-recognizing proteins [4].

Trbp111 is a 111-amino acid polypeptide initially identified in *Aquifex aeolicus* that nonspecifically binds tRNA species. It is homologous to other bacterial proteins and to domains of the bacterial methionyl-tRNA synthetases (MetRSs) and phenylalanyl-tRNA synthetases (PheRSs) [5–8]. Crystal structures of Trbp111 and its counterpart in *Escherichia coli* revealed a dimeric organization and the presence of a conventional oligonucleotide/oligosaccharide binding (OB)-fold in each monomer (Fig. 1B) [9]. Trbp111 contains a canonical OB-fold structure, with five β-strands assembled into a β-barrel and a small α-helix (helix α2) inserted between β-strands β3 and β4. The dimerization interface is assembled from helix α1 and β-strands β6-β8 from each monomer [10]. The tRNA binding site in Trbp111 is formed in part by the dimer interface. Consistently, Trbp111 experimentally binds tRNA in a 2 : 1 ratio [4]. A model of the complex between Trbp111 and tRNA suggested that Trbp111 interacts with the outside corner of the L-shaped tRNA formed by the D- and T-loops; this model explains Trbp111's lack of specificity for definitive tRNA isoacceptors. Analysis of the Trbp111 footprint on tRNA showed that the dimer also protects surrounding nucleotides located in the T-, D-, and anticodon arms beyond the outside corner of the L-shaped tRNA [11]. The apparent dissociation constant between Trbp111 and tRNA ($K_d$) is 32 nM. Extensive mutagenesis identified six amino acids (K33, E45, N71, R75, S82, and M85) important for tRNA binding, with mutations at S82 and M85 having the greatest effect (Fig. 1B) [9]. A single tRNA can concomitantly bind to Trbp111 and its cognate aminoacyl-tRNA synthetase (aaRS), effectively sandwiched between the two proteins with the aaRS binding the acceptor stem and, in many cases, the anticodon stem, while Trbp111 binds the tRNA elbow. Formation of this complex enhances the rate of aminoacylation, indicating that the tRNA tertiary structure is intact [11] and suggesting that Trbp111 can act as a cofactor for aminoacylation, either by stabilizing the fold of the bound tRNA or by inducing a conformational adjustment.

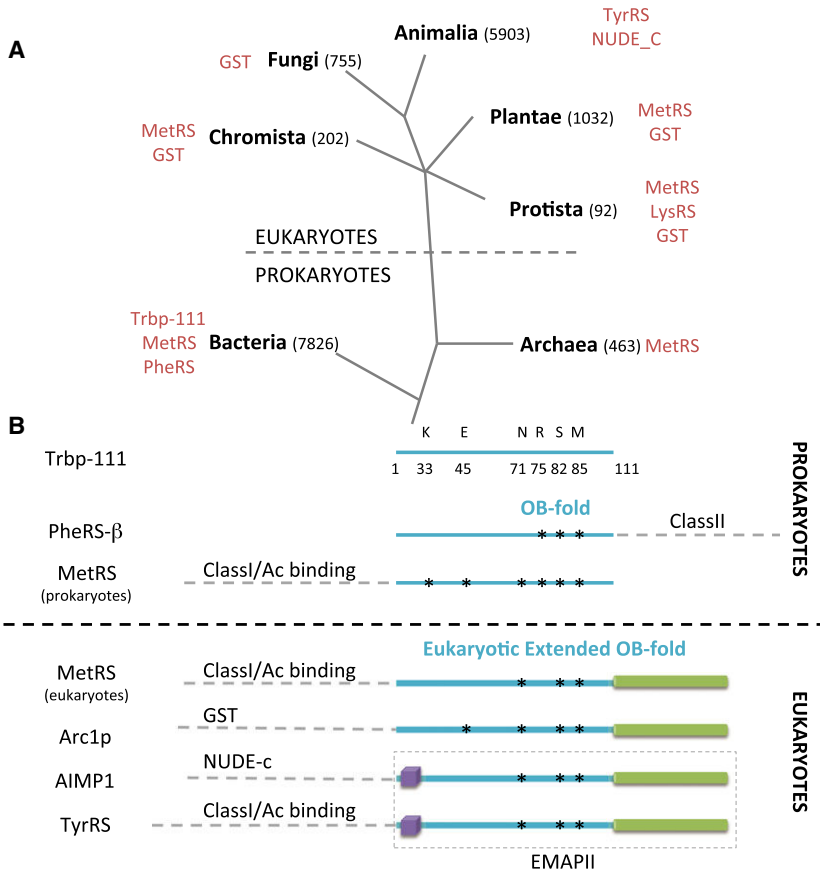



**Fig. 2.** Evolution and organization of OB-fold-containing proteins. (A) Phylogenetic tree based on the classification by [53], indicating N-terminal domains fused to the OB-fold, which were identified and studied in prokaryotes (bacteria and archaea) and eukarya. The numbers of genomes sequenced in each taxon are indicated in brackets. (B) Modular organization of OB-fold-containing proteins. The OB-fold domain is shown in blue; the extended C-terminal domain in green; the migration heptapeptide signal in purple; and the conservation of amino acids involved in tRNA binding is indicated with black stars.

Since the discovery of Trbp111, a free-standing OB-fold protein, a number of other proteins and enzymes have been identified that contain OB-fold domains. Figure 2 shows representations of well-known protein fusions containing a Trbp111 OB-fold domain; evolutionary fusion events have occurred with aminoacyl-tRNA synthetases (aaRSs), glutathione S-transferase (GST), and NUDE_C modules. Binding of an OB-fold domain to a tRNA elbow can significantly increase the binding affinity of the fused protein and also ensure that the tRNA structure retains its canonical L-shape, offering an explanation for the unusual distribution of this domain through evolution. Yet, some remarkable examples of OB-fold-containing proteins perform very different functions. Here, we give an overview of the 'OB-fold family', that is to say any protein with a domain homologous to Trbp111. We describe the distribution of these fusion proteins in the tree of life, their ability to bind the tRNA elbow structure, and their alternative functions. We observed that OB-fold domains are fused to many different proteins, like the well-characterized MetRS and GST proteins, but also to unexpected aaRSs and the Yip1 membrane protein, especially in unicellular pathogenic eukaryotes.

## Inventory of recurrent OB-fold fusions

As an independent and minimal module, Trbp111 exists only in bacteria. However, OB-fold fusions are found in the three domains of life (Fig. 2A). With respect to this review, we will focus on three classes of enzymes, the aaRSs, GST, and NUDE_C, which contain examples of acquired OB-fold domains. In these cases, the OB-fold can be appended N-terminally (e.g., the β-subunit of PheRS), as an independent C-terminal domain (e.g., prokaryotic MetRS), or followed by a conserved C-terminal motif (e.g., eukaryotic TyrRS) (Fig. 2B). The occurrence of this C-terminal extension is specific to eukaryotic OB-fold-containing proteins. It is ~ 100–200 amino acids long, and structural analysis showed that, together, the monomeric OB-fold and its extended C-terminal sequence recapitulate a portion of the Trbp111 dimeric structure (Fig. 1B). This class of OB-fold binds tRNA species in a 1 : 1 ratio [12].

In many organisms, four aaRSs contain an OB-fold domain, regardless of their aaRS classification (Fig. 2): bacterial PheRS (class II), MetRS (class I) in all three domains of life, tyrosyl-tRNA synthetase (TyrRS, class I) only in vertebrates, and lysyl-tRNA synthetase

(LysRS, class II) in amoeba. The aaRS family of enzymes is an essential component of the translation apparatus. They work hand in hand with tRNA species, by catalyzing the first reaction step in the synthesis of proteins, tRNA aminoacylation [13]. Most aaRSs primarily recognize nonelbow features of their cognate tRNA substrates, like the anticodon domain and/or the acceptor stem of the molecule (Fig. 1A) [14]. Thus, the presence of a general tRNA-binding domain in the vicinity of an aaRS catalytic site, like an OB-fold fusion, could facilitate amino acid charging by increasing the local concentration of the tRNA to improve efficiency and maintain optimal aminoacylation levels without compromising specificity [15]. We will discuss these enzymes in more detail later in this review.

OB-fold fusions to GST and NUDE_C proteins are usually involved in protein–protein interactions [16]. GST fusions are found in plants and unicellular eukaryotes, whereas NUDE_C fusions are present only in vertebrates (Fig. 2A). Both fusions allow the formation of protein complexes with aaRSs, which contributes to the efficiency of translation by preventing the diffusion of charged tRNA into the cytoplasm [17]. In the next few paragraphs, we focus on what is currently known about NUDE_C proteins with appended OB-fold domains. GST fusions are described later in this review in the section devoted to unicellular pathogenic eukaryotes.

In vertebrates, nine aaRSs are located in a multisynthetase complex (MARS) with three auxiliary proteins called aminoacyl-tRNA synthetase complex interacting multifunctional proteins 1, 2, and 3 (AIMP1, AIMP2, and AIMP3) [18,19]. The C-terminal OB-fold of AIMP1 binds tRNA species and contributes to the efficiency of translation [20]. AIMP1 dimerization is sustained by its NUDE_C N-terminal domain, which contains two putative coiled-coil motifs. The first is responsible for this protein's association with the corresponding motif in AIMP2, whereas the second is required for interactions with the glutaminyl-tRNA synthetase (GlnRS) and arginyl-tRNA synthetase (ArgRS) in the MARS complex [21]. AIMP1's role in delivering $tRNA^{Arg}$ to ArgRS is still controversial [22].

The NUDE_C module of AIMP1 is found exclusively in vertebrates, and its proteolytic cleavage unmasks the emergence of an additional function for its OB-fold. AIMP1 is the precursor of the endothelial monocyte activating polypeptide II (EMAPII) [20]. As described above, full-length AIMP1 is part of MARS and binds tRNA species and aaRSs. When AIMP1 is cleaved by caspase 7 [23], EMAPII is produced. EMAPII comprises only the extended OB-fold domain and it dissociates from the MARS complex [24]. EMAPII fails to bind tRNA species (apparent $K_D$ of 40 μM). Instead, it possesses various cytokine activities critical to cell signaling and developmental pathways. In humans, EMAPII has a broad spectrum of action, including the induction of apoptosis and migration in endothelial cells, and migration of inflammation to related cell types, like macrophages and monocytes [25,26].

### The OB-fold domain in unicellular pathogenic eukaryotes

EuPathDB is an integrated database covering eukaryotic pathogen genomes [27]. They represent a diverse group of eukaryotic single-celled organisms, often with only superficial similarities. These pathogens are responsible for extensive morbidity and mortality worldwide and cause a variety of diseases in plants and animals. In humans, prevalent eukaryotic pathogens are divided into (a) intestinal pathogens such as *Entamoeba histolytica*, *Giardia intestinalis*, and *Cryptosporidium parvum*; (b) urogenital pathogens such as *Trichomonas vaginalis*; and (c) parasites that infect blood and tissues, such as *Toxoplasma gondii* and *Plasmodium falciparum*. OB-fold domains are found in many different proteins in these pathogens and efforts to elucidate their functional importance demonstrate that these domains play a variety of different roles with respect to tRNA biology.

## tRNA binding potential of OB-folds in unicellular eukaryotic pathogens

### C-terminal extended OB-folds (OB-C)

In eukaryotes, a single, monomeric protein partially mimics the homodimeric structure of Trbp111, the major difference being the addition of the C-terminal extended OB-fold domain [12] (Figs 1B and 2B). In unicellular eukaryotes, many isolated OB-fold domains are characterized by this prolonged C-terminal sequence, but are deprived of any fused N-terminal structural module (Fig. 3). These are designated as OB-C proteins in this review; examples are found in most eukaryotic pathogens, except the *Apicomplexa* and nearly all fungi (Table 1). OB-C proteins are typically between 200 and 300 amino acids long. One OB-C homolog of Trbp111 has been studied in the *kinetoplastida Trypanosoma* [28]. The *Kinetoplastida* group includes a number of parasites that cause serious diseases in humans and other animals, including sleeping sickness and Chagas disease, caused by

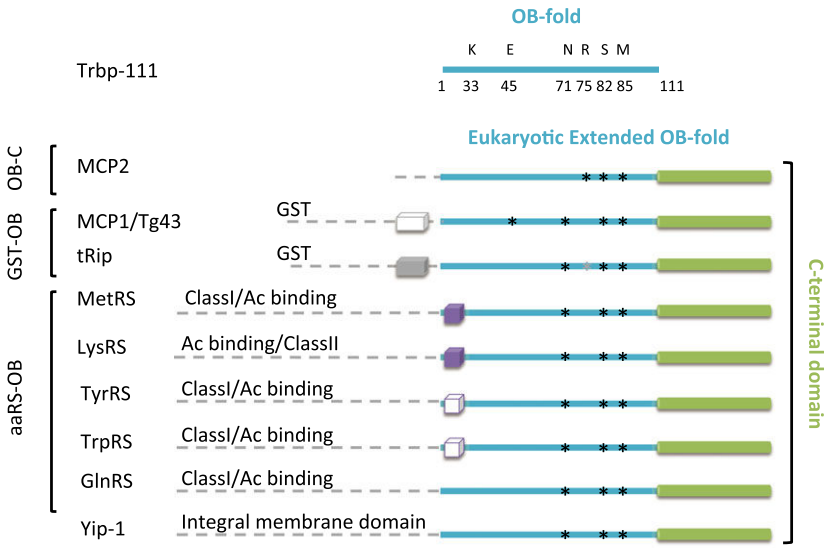



**Fig. 3.** Modular organization of OB-fold-containing proteins from pathogenic unicellular eukaryotes. The same color code was used as in Fig. 2B. tRip contains a K instead of R75 (indicated with a gray star).

*Trypanosoma* species, and leishmaniasis, caused by *Leishmania* species; various *kinetoplastida* forms are also found in soil and aquatic environments.

In *T. brucei*, the OB-C MARS complex-associated protein 2 (MCP2) is a component of the *Trypanosoma* MARS complex, which also encompasses six aaRSs and two other MCP proteins (MCP1 and MCP3). MCP2 contributes to efficient tRNA aminoacylation [28]. Recombinant MCP2 forms a homodimer and shows binding selectivity to tRNA species over other RNA species *in vitro*. MCP2 depletion from the cell resulted in slow parasite growth and decreased infectivity in mouse, presumably reflecting its role in increasing the efficiency of tRNA aminoacylation.

Free-standing OB-C proteins are the only OB-fold proteins found in *Microsporidiae* and *Diplomonadida*, both of which cause opportunistic intestinal infection in animals. In other pathogens, up to three copies of OB-C genes can be found (e.g., in *Trichomonas vaginalis*) and/or in combination with GST- or aaRS-OB fusions.

## GST-OB fusions

### In fungi

Fusions between the C-terminal extended OB-fold and a GST-like domain (Fig. 2A) are the most abundant OB fusions found in eukaryotic pathogens. The GST fusion is the only source of OB-fold in most fungi. A well-known example is the aaRS cofactor 1 (Arc1p) protein from the fungi *S. cerevisiae*. *S. cerevisiae* is now regarded as an opportunistic pathogen, albeit of low virulence. Existing data indicate that Arc1p shares a similar mode of interaction with tRNA species as Trbp111. However, the selection and binding of tRNA species are facilitated by two additional modules: the extended C-terminal module (like in OB-C proteins), which recognizes the conserved 'elbow' structure of the tRNA, and an N-terminal lysine-rich domain, which binds RNA nonspecifically [29]. The Arc1p/tRNA complex was studied by cross-linking and gel retardation experiments and showed a binding affinity corresponding to a $K_D$ value of about 10 nM; this tight affinity requires both the C-terminal extended OB-fold and the lysine-rich domains [29].

The GST-like module does not possess enzymatic activity, but is used for the assembly of an aaRS complex [16]. In *S. cerevisiae*, this domain interacts with the N-terminal domains of glutamyl-tRNA synthetase (GluRS) and MetRS simultaneously [30]. Both GluRS and MetRS gain a catalytic benefit by associating with Arc1p, as shown in *in vitro* reconstitution and enzymatic analyses of the corresponding complexes [31,32]. Thus, the Arc1p OB-fold domain is crucial for facilitating the binding of $tRNA^{Glu}$ and $tRNA^{Met}$ to their respective synthetases. Interestingly, both $tRNA^{Met}$ and $tRNA^{Glu}$ can bind to the Arc1p/MetRS/GluRS complex, yet binding of the one tRNA does not influence the binding efficiency of the other tRNA: Each tRNA binds to this complex independently of the other, without synergy or interference. However, it is not known whether both tRNA species can bind to Arc1p simultaneously or whether they bind one at a time [29,31,33]. Moreover, in fermenting yeasts, Arc1p anchors GluRS and MetRS in the cytosol [34]. When yeast cells switch to respiration, the Arc1p/MetRS/GluRS triad dissociates. Then, MetRS and GluRS

**Table 1.** List of OB-fold-containing proteins identified in EuPathDB genomes.

| DATABANK | | SPECIES | OB-C | GST-OB | aaRSs-OB | |
|---|---|---|---|---|---|---|
| **DIPLOMONADIDA** | *GIARDIA* | *G. intestinalis* | 1 | | | |
| GiardaDB | *SPIRONUCLEUS* | *S. salmonicida* | 1 | | | |
| **MICROSPORIDIUM** | *NOSEMA* | *N. ceranae* | 1 | | | |
| MicrosporidiaDB | | *N. bombycis* | 1 | | | |
| | *MITOSPORIDIUM* | *M. daphniae* | 1 | | | |
| | *TRACHIPLEISTOPHORA* | *T. hominis* | 1 | | | |
| | *ENCEPALITOZOON* | *E. intestinalis* | 1 | | | |
| | | *E. hellem* | 1 | | | |
| | | *E. romaleae* | 1 | | | |
| | | *E. cuniculi* | 1 | | | |
| | *ENTEROCYTOZOON* | *E. bieneusi* | 1 | | | |
| | *SPRAGUEA* | *S. lophii* | | 1 | | |
| | *VAVRAIA* | *V. culicis* | | 1 | | |
| | *VITTAFORMA* | *V. corneae* | 1 | | | |
| | *ANNCALIIA* | *A. algerae* | 1 | | | |
| | *ORDOSPORA* | *O. colligata* | 1 | | | |
| | *EDHAZARDIA* | *E. aedis* | 1 | | | |
| | *NEMATOCIDA* | *N. parisii* | 1 | | | |
| **FUNGI** | *Agaricomycetes* | | | 1 | | |
| FungiDB | *Blastocladiomycetes* | | | 1 | | |
| | *Chytridiomycetes* | | | | | |
| | *Eurotiomycetes* | | | 1 | | |
| | *Leotiomycetes* | | | 1 | | |
| | *Pneumocystidomycetes* | | | | | |
| | *Pucciniomycetes* | | | 1 | | |
| | *Schizosaccharomycetes* | | | 1 | | |
| | *Sordariomycetes* | | | 1 | | |
| | *Tremellomycetes* | | | 1 | | |
| | *Ustilaginomycetes* | | | 1 | | |
| | *Zygomycetes* | | | 1 | | |
| | *Crithidia* | | | | | |
| | *Endotrypanum* | | | | | |
| | *Saccharomycetes* | | | 1 | | |
| | | *Y. lipolytica* | 1 | 1 | | |
| | *Zygomycetes* | *M. circinelloides* | 2 | | | |
| | *Zygomycetes* | *P. blakesleeanus* | 2 | | | |
| | *Blastocladiomycetes* | *A. macrogynus* | 2 | | | |
| | *Chytridiomycetes* | *S. punctatus* | | 1 | | TyrRS GlnRS |
| | | *B. dendrobatidis* | | 1 | | TyrRS GlnRS |
| **OOMYCETES** | *SAPROLEGNIA* | *S. parasitica* | 2 | | MetRS | |
| | | *S. diclina* | 2 | | MetRS | |
| | *HYALOPERONOSPORA* | *H. arabidopsidis* | 2 | | MetRS | |
| | *PHYTOPHTHORA* | *P. capsici* | 1 | | MetRS | |
| | | *P. infestans* | 1 | | MetRS | |
| | | *P. cinnamomi* | 1 | | MetRS | |
| | | *P. parasitica* | 1 | | MetRS | |
| | | *P. ramorum* | 1 | | MetRS | |
| | | *P. sojae* | 1 | | MetRS | |
| | *APHANOMYCES* | *A. invadans* | 1 | | MetRS | |
| | | *A. astaci* | 1 | | MetRS | |
| | *PYTHIUM* | *P. aphanidermatum* | 1 | | MetRS | Yip1 |
| | | *P. irregulare* | 2 | | MetRS | |
| | | *P. ultimum* | 2 | | MetRS | |
| | | *P. arrhenomanes* | 1 | | MetRS | |
| | | *P. vexans* | 1 | | MetRS | |
| | | *P. iwayamai* | 2 | | | |
| | *ALBUGO* | *A. laibachii* | 2 | | MetRS | |
| | | *A. candida* | 2 | | MetRS | |
| **TRICHOMONADIDA** | *TRICHOMONAS* | *T. vaginalis* | 3 | 1 | | |
| TrichDB | | | | | | |

| DATABANK | | SPECIES | OB-C | GST-OB | aaRSs-OB | |
|---|---|---|---|---|---|---|
| **AMOEBOZOA** | *NAEGLERIA* | *N. floweri* | 2 | 1 | MetRS | |
| AmoebaDB | *ACANTHAMOEBA* | *A. castellanii* | 1 | 1 | | TyrRS |
| | *ENTAMOEBA* | *E. invadens* | | | MetRS | LysRS |
| | | *E. histolitica* | | | MetRS | LysRS |
| | | *E. moshkovskii* | | | MetRS | LysRS |
| | | *E. dispar* | | | MetRS | LysRS |
| **CHROMERIDA** | *VITRELLA* | *V. brassicaformis* | 1 | 1 | MetRS* | TrpRS |
| cryptoDB | *CHROMERA* | *C. velia* | 2 | 1 | MetRS | |
| **APICOMPLEXA** | *BABESIA* | *B. microti* | | 1 | | |
| PiroplasmaDB | | *B. bovis* | | 1 | | |
| | | *B. bigemina* | | 1 | | |
| | *THELEIRIA* | *T. equi* | | 1 | | |
| | | *T. parva* | | 1 | | |
| | | *T. orientalis* | | 1 | | |
| | | *T. annulata* | | 1 | | |
| | *CYTAUXZOON* | *C. felis* | | 1 | | |
| PlasmoDB | *PLASMODIUM* | *P. chabaudi* | | 1 | | |
| | | *P. yoelii* | | 1 | | |
| | | *P. berghei* | | 1 | | |
| | | *P. knowlesi* | | 1 | | |
| | | *P. reichenowi* | | 1 | | |
| | | *P. falciparum* | | 1 | | |
| | | *P. cynomolgi* | | 1 | | |
| | | *P. vivax* | | 1 | | |
| ToxoDB | *EIMERIA* | *E. praecox* | | 1 | | |
| | | *E. falciformis* | | 1 | | |
| | | *E. acervulina* | | 1 | | |
| | | *E. mitis* | | 1 | | |
| | | *E. necatrix* | | 1 | | |
| | | *E. tenella* | | 1 | | |
| | *CHRYPTOSPORIDIUM* | *C. Hominis* | | 1 | | |
| | | *C. parvum* | | 1 | | |
| | | *C. muris* | | 1 | | |
| | *GREGARINA* | *G. niphandrodes* | 1 | | | |
| | *NEOSPORA* | *N. caninum* | | 1 | | |
| | *SARCOCYSTIS* | *S. neurona* | | 1 | | |
| | *TOXOPLASMA* | *T. gondii* | | 1 | | |
| | *HAMMONDIA* | *H. hammondi* | | 1 | MetRS* | |
| **KINETOPLASTIDA** | *TRYPANOSOMA* | *T. brucei* | 1 | 1 | MetRS | |
| TriTrypDB | | *T. evansi* | 1 | 1 | MetRS* | |
| | | *T. grayi* | 1 | 1 | MetRS* | |
| | | *T. cruzi* | 1 | 1 | | |
| | | *T. rangeli* | 1 | 1 | | |
| | | *T. vivax* | 1 | 1 | | |
| | | *T. congolense* | 1 | | | |
| | *LEPTOMONAS* | *L. seymouri* | 1 | 1 | | |
| | | *L. pyrrhocoris* | 1 | 1 | | |
| | *LEISHMANIA* | *L. panamensis* | 1 | 1 | | |
| | | *L. braziliensis* | 1 | 1 | | |
| | | *L. aethiopica* | 1 | 1 | | |
| | | *L. tropica* | 1 | 1 | | |
| | | *L. donovani* | 1 | 1 | | |
| | | *L. infantum* | 1 | 1 | | |
| | | *L. arabica* | 1 | 1 | | |
| | | *L. enriettii* | 1 | 1 | | |
| | | *L. gerbilli* | 1 | 1 | | |
| | | *L. major (3X)* | 1 | 1 | | |
| | | *L. mexicana* | 1 | 1 | | |
| | | *L. tarentolae* | 1 | 1 | | |
| | | *L. turanica* | | 1 | | |
| | *CRITHIDIA* | *C. fasciculata* | 1 | 1 | | |
| | *ENDOTRYPANUM* | *E. monterogeii* | 1 | 1 | | |

The sequences of *S. cerevisiae* Arc1p and *A. aeolicus* Trbp111 were BLASTed [55] against each EuPathDB databank. OB-C means that no defined structural domain was identified at the OB-fold N-terminal extremity. GST-OB and aaRS-OB mean that the OB-fold is fused at the C-terminal extremity of GST or of an aaRS, respectively. Proteins that are indicated in red correspond to newly discovered aaRS fusions. Species are highlighted with different colors depending on the number of genes coding for OB-fold-containing proteins in their genomes: one copy (beige), two copies (green), three copies (light blue), and four copies (dark blue).

relocalize into the nucleus and the mitochondria, respectively, where they coordinate expression of the different components of the respiratory chain [35]. It has also been proposed that Arc1p could act as an adapter in nucleocytoplasmic trafficking of tRNA [32].

## In *Apicomplexa*

The GST-OB-fold is also present in *Apicomplexa*. Many of the apicomplexan parasites are important pathogens of human and domestic animals and lead to devastating infections, such as malaria and toxoplasmosis. For the purposes of this review, we have divided the *Apicomplexa* parasites into two groups depending on their hosts (Table 1). The first group, including *Cryptosporidium*, *Neospora*, *Sarcocystis*, *Toxoplasma,* and *Hammondia*, develops in a unique vertebrate host, and the second group, including *Babesia*, *Theileria*, *Cytauxzoon*, *Plasmodium*, *Eimeria,* and *Gregarina*, develops in both insect and vertebrate hosts.

### *Toxoplasma*

The first group of *Apicomplexa*, with the exception of *Cryptosporidium*, contains both GST- and aaRS-OB fusions. The GST-OB found in *Toxoplasma gondii* was named Tg-p43. As a dimer, it belongs to a MARS complex that also contains four aaRSs. Surprisingly,

when the corresponding Tg-p43 gene was deleted, the parasite was viable without any obvious defects in culture and with no differences in infectivity or lethality when injected into mice and compared with the parental strain [36]. This observation suggests that Tg-p43 is either nonessential or only essential under specific conditions that have not yet been identified or tested (e.g., stress conditions).

*Plasmodium*

The GST-OB fusion is the only OB-fold found in the second group of *Apicomplexa*. *Plasmodium*, which causes malaria, is representative of this group and they require two different hosts (insects and mammals). In *Plasmodium*, the tRNA import protein tRip is an Arc1p homolog [37]. *In vitro*, tRip binds human tRNA species with high affinity ($K_D$ = 5–10 nM) and recognizes the elbow of the molecule *via* its C-terminal OB-fold. tRip contains a transmembrane domain and is located at the surface of the parasite with its OB-fold domain exposed on the outside of the parasite. The tRip transmembrane domain, responsible for this peripheral localization, was predicted only in *Apicomplexa* genomes. tRip mediates host tRNA uptake specifically into the parasite sporozoites. *Plasmodium* sporozoites, which accumulate in *Anopheles* mosquito salivary glands, are the infectious form of the parasite and are transmitted to the vertebrate host. In the absence of tRip, sporozoite development is not detectibly disturbed, but translation and parasite infectivity is diminished at the blood stage. Because mature red blood cells do not contain tRNA species, the authors proposed that the observed phenotype could be the result of either a delayed effect of imported tRNA species at the sporozoite stage, or tRNA import in reticulocytes (precursors of red blood cells that contain tRNA species).

*Kinetoplastida*

In *kinetoplastida*, the GST-OB fusion MCP1 coexists with the MCP2 OB-C. MCP1 is found in the same MARS complex as MCP2, but has not been further studied [28].

### AaRS-OB fusions

In unicellular pathogens, MetRS is the most common aaRS-OB protein. It is present in *Amoeba*, *Chromerida*, *Oomycetes,* and some *Apicomplexa* (group 1, with the exception of *Cryptosporidium*). MetRS and lysyl-tRNA synthetase (LysRS) from *E. histolytica* share the same OB-fold [38]. *E. histolytica* is the most devastating cause of dysenteric human death in the world [39]. As with the OB-C and GST-OB fusions, this tRNA-binding OB-fold module is followed by an extended C-terminal sequence. Disruption of the *E. histolytica* C-terminal OB-fold appended to LysRS reduced aminoacylation efficiency by 40-fold, with effects on both the $K_M$ and the $k_{cat}$ of the reaction. This result suggests that the OB-fold binds tRNA species and resembles other eukaryotic MetRSs that also contain this C-terminal OB-fold. Functional studies of the *O. sativa* and *C. elegans* MetRSs led to the conclusion that MetRS OB-fold domains bind tRNA species as monomers (1 : 1 ratio) and provide the enzymes with better catalytic efficiency [40,41]. In *O. sativa*, the OB-fold produces a stable complex with tRNA ($K_D$ = 120 nM) and facilitates aminoacylation (10-fold decrease in $K_M$ for $tRNA^{Met}$ aminoacylation). Interestingly, cognate and noncognate tRNA species bound to MetRS with similar affinities, but only $tRNA^{Met}$ was aminoacylated. Therefore, the general RNA binding capacity of the OB-fold did not interfere with the tRNA aminoacylation selectivity of MetRS [40].

*In vivo*, *E. histolytica* LysRS and MetRS can be proteolytically cleaved to release their C-terminal OB-fold domains, both of which include a heptapeptide migration motif (RVGKIVE). When released, the OB-fold protein attracts human endothelial cells, but not monocytes. The authors proposed that LysRS and MetRS OB-fold domains use this chemotactic activity to modulate the host immune response and to improve parasite survival and fitness [38].

## Other aaRS-OB fusions specific for unicellular eukaryotes

While preparing this review, three unexpected aaRS-OB fusions were identified, corresponding to TrpRS, GlnRS, and TyrRS fusions (Fig. 3, Table 1). In particular, the presence of a C-terminal OB-fold in TyrRSs in a unicellular eukaryote is notable because this fusion event had previously only been observed in arthropods and vertebrates. In these organisms, removing the OB-fold does not change aminoacylation activity of the TyrRS [42], suggesting that the OB-fold does not bind tRNA species or at least does not synergistically enhance tRNA binding to TyrRS in a catalytically definable way. However, in vertebrates, this OB-fold is a structural and functional EMAPII homolog with similar cytokine functions as observed with the AIMP1 OB-fold. The TyrRS and AIMP1 EMAPII domains also display heptapeptide migration motifs [42]: RVGKIIT in AIMP1 and RIGCIIT in TyrRS

(Figs 1B, 2B, and 4). This motif slightly differs from the corresponding sequence in *Entamoeba* LysRS and MetRS (KVGKIIE) (Fig. 4). These sequence variations are enough to promote different chemotactic effects: On the one hand, *Entamoeba* MetRS and LysRS OB-fold domains, when processed to discreet proteins, directly attract human endothelial cells, which subsequently generate a chemoattractant signal for monocytes [38]. On the other hand, AIMP1 and TyrRS-OB-folds induce directional migration of mononuclear phagocytes and polymorphonuclear leukocytes and activate endothelial cells by stimulating leukocyte adherence. AIMP and TyrRS also inhibit angiogenesis, play a role in regulating tumor vascular growth, and up-regulate the production of TNF, tissue factors and the release of myeloperoxidase from polymorphonuclear leukocytes [43]. Truncation of EMAPII from TyrRS exposes an ELR cytokine tripeptide motif, which is otherwise buried in the catalytic domain [44]. This ELR sequence stimulates angiogenesis and evolutionarily begins to appear in insect TyrRS, simultaneously with the addition of the C-terminal OB-fold domain.

The presence of TyrRS-OB in isolated unicellular eukaryotes raises many questions. The amoeba *Acanthamoeba castellanii* and the fungi *Spizellomyces punctatus* and *Batrachochytrium dendrobatidi* are three pathogens with very distinct hosts leading to different diseases. Amoeba *A. castellanii* is an opportunistic human parasite responsible for a fatal form of encephalitis [45]. *S. punctatus* is an important member of soil microbial communities influencing plant production, both adversely and beneficially [46]. And, *B. dendrobatidis* is an aquatic fungus, which causes

| Species | Fusion Proteins | Heptapeptide |
|---|---|---|
| CONSENSUS | | xIGx(I/V)(T/E) |
| ***Mus musculus*** | AIMP1 | RIGRIVT |
| | | RIGRIIT |
| | | AIGRIVT |
| | | (RIGRAVT) |
| ***Homo sapiens*** | AIMP1 | RIGCIIT |
| | TyrRS | RVGKIIT |
| ***E. invadens*** | LysRS | RVGEIVE |
| | MetRS | RVGQIVE |
| ***E. histolitica*** | LysRS | RVGKIVE |
| | MetRS | RVGKIVE |
| ***E. moshkovskii*** | LysRS | KVGKIVE |
| | MetRS | KVGKIVE |
| ***E. dispar*** | LysRS | KVGKIIE |
| | MetRS | RVGKIVE |
| ***V. brassicaformis*** | TrpRS | RVGRILE |
| ***N. floweri*** | MetRS | RVGQIET |
| ***A. castellanii*** | TyrRS | KVGRIVD |
| ***S. punctatus*** | TyrRS | RVGRIVD |
| | GlnRS | CVGRILD |
| ***B. dendrobatidis*** | TyrRS | RVGKVIE |
| | GlnRS | CVGRILD |
| ***G. prolifera*** | TyrRS | RVGRIVS |
| | GlnRS | RVGKIKS |
| ***M. elongata*** | TyrRS | RVARILE |
| ***M. verticillata*** | TyrRS | RVARILE |
| ***R. allomycis*** | TyrRS | VVGEIKE |
| ***R. irregularis*** | TyrRS | RVAKIIK |
| ***T. deformans*** | TyrRS | RVGLIED |

**Fig. 4.** Sequences of the migration heptapeptide signals and possible conservation in eukaryotic pathogen aaRS-OB fusions. The mouse migration heptapeptide sequence and related mutants were retrieved from [54]. The consensus sequence was determined based on this study and the sequences found in human AIMP1 and TyrRS and in amoeba MetRS and LysRS. Putative migration heptapeptides from unicellular eukaryotes were identified by alignment of their OB-fold with human AIMP1 and TyrRS and amoeba MetRSs and LysRS OB-folds using ClustalW (http://www.ch.embnet.org/software/ClustalW.html).

fatal infectious disease in amphibians by blocking their respiratory organs [47]. TyrRS-OB was also identified in *Taphrina deformans* (a plant pathogen absent from EuPathDB), and five other nonpathogenic fungi genomes (NCBI Blast): *Rozella allomycis*, *Mortierella verticillata*, *Mortierella elongate*, *Gonapodya prolifera*, and *Rhizophagus irregularis*, suggesting that the presence of a TyrRS-OB is not correlated with the pathogenicity of the fungi.

The genomes of *S. punctatus*, *B. dendrobatidis* (Table 1), and *Gonapodya proliferam* contain not only a TyrRS-OB but also a GlnRS-OB, whereas *V. brassicaformis* seems to be the only organism containing a TrpRS-OB. GlnRS-OB and TrpRS-OB fusions do not have homologs in other organisms. These proteins have not yet been characterized; thus, the putative role played by their respective OB-fold is enigmatic.

## A unique Yip1-OB fusion

Yip1 family members (Fig. 3) are transmembrane proteins that participate in the transport pathway that serves to translocate proteins from one membrane location to another. Yip1 interacts with GTPases and is involved in endoplasmic reticulum (ER) to Golgi membrane transport and remains under study [48]. Yip1 proteins contain four transmembrane α-helices and the N terminus is oriented in the cytosol. Thus, in the plant pathogen *Pythium aphanidermatum* (Table 1), the C-terminal OB-fold of the unique Yip1-OB fusion should be located on the cytosolic side of the ER. The function of this OB-fold domain remains unknown.

## Conclusions

All eukaryotic unicellular pathogen genomes found in EuPathDB contain at least one copy of an OB-fold domain or protein and show the propensity to add new modules to this OB-fold via evolutionary fusion events. Most of these N-terminal extensions participate in well-known tRNA biology (e.g., tRNA aminoacylation). In many cases, the presence of this nonspecific tRNA-binding domain in the neighborhood of an aaRS catalytic site increases the affinity of the tRNA to the aaRS to enhance tRNA aminoacylation [38]. Likewise, GST fusions, which allow the formation of protein complexes with aaRSs, contribute also to the efficiency of tRNA charging (e.g., in *S. cerevisiae*) [31,32]. However, new functions have arisen in unicellular eukaryotes. For example, in *Plasmodium*, the GST fusion in tRip is localized in the parasite membrane and is involved in tRNA trafficking [37]. The Yip1-OB fusion present in *P. aphanidermatum* is another example of a possible membrane protein with an OB-fold tRNA-binding domain, of unknown function in this case. Furthermore, many of these OB-fold proteins are isolated polypeptides with no significant N-terminal domain. In this circumstance, when the OB-fold is separate from an aminoacylation context (e.g., not attached to an aaRS or part of a MARS), one wonders about the selectivity of this domain toward tRNA sequences. Deinert and collaborators showed that under competition conditions, the OB-fold of Arc1p binds preferentially to a subset of tRNA species, not only to $tRNA^{Glu}$ and $tRNA^{Met}$, but also to $tRNA^{Phe}$, $tRNA^{Lys}$, and $tRNA^{Arg}$ [31]. These tRNA species display scattered G–C elements that may act as positive determinants for the interaction with the OB-fold. This selectivity is reminiscent of the initial identification of Arc1p (then named G4p1) as a protein with affinity for DNA and RNA G-quadruplexes [49]. Indeed, Arc1p binds DNA and RNA quadruplex structures that are found in gene promoters, in 5′ and 3′ untranslated regions, and especially in subtelomeric regions [50].

While retaining the amino acids responsible for tRNA binding (especially N71, S82, and M85), do the OB-fold domains serve specific roles in unicellular pathogenic eukaryotes? The presence of heptapeptide migration motifs is consistent with the possibility that some OB-fold domains are involved in interactions with host target cells. No accurate sequence homology with the consensus sequence XIGXI(V/I)(T/E) (Fig. 4) is found in pathogenic eukaryotic unicellular proteins, except in the OB-fold domains of *Entamoeba* LysRS and MetRS; when proteolyzed, these OB-folds are chemoattractants for human endothelial cells, direct evidence of communication between OB-fold proteins and host target cells [38]. In *Chromerida* (algal ancestor of malaria parasites), *Oomicetes* (plant parasites), *Spironucleus salmonicida* (fish parasite), and some *Microsporidia* (infectious agents against specific insects and vertebrates), some OB-C and GST-OB sequences differ from this heptapeptide migration consensus sequence by only one amino acid. For example, in *Trichomonas vaginalis* (the parasite responsible for trichomoniasis), the four OB-fold-containing proteins encoded in the genome display conserved heptapeptide sequences where only the amino acid at position 7 is different from the abovementioned consensus sequences (RVGKIV**K**, RVGKIV**S**, RVGKII**H**, and RVGKII**Q**). Likewise, most of the aaRS-OB fusions identified (Table 1) show only a few changes in this migration heptapeptide sequence (Fig. 4). Like in *Entamoeba* MetRS and LysRS, one can imagine that

these differences may influence the specificity of host cell attraction. Further studies will be necessary to precisely determine potential structure–function relationships of these newly identified OB-fold domains.

Moreover, in *S. punctatus*, a plant pathogen, TyrRS not only contains an RVGRIVD heptapeptide sequence in its OB-fold domain, but also an ELR sequence that aligns with the ELR sequence exclusively found in insect and vertebrate TyrRSs. This is an unusual and currently inexplicable coincidence when one considers that this parasite infects plants, but not vertebrates. The discovery in pathogens of fusions homologous to the vertebrate TyrRS-OB raises interesting questions concerning its evolutionary history and functional role.

The accumulation of new aaRS-OB fusions in pathogenic unicellular eukaryotes highlights the largely uncharacterized, biological complexity of this domain. The emergence of idiosyncratic aaRS-OB fusions, the OB-fold's neglected affinity for G-quadruplex structures, as well as the conservation of potential cytokine sequences, suggest their involvement in a wide range of as yet undiscovered biological functions.

## Author contributions

DK, MC, AT-D, TH, and MF participated in drafting the review.

## References


1 Kim S (1975) Symmetry recognition hypothesis model for tRNA binding to aminoacyl-tRNA synthetase. *Nature* **256**, 679–681.

2 Westhof E, Dumas P and Moras D (1988) Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals. *Acta Crystallogr A* **44** (Pt 2), 112–123.

3 Zhang J and Ferre-D'Amare AR (2016) The tRNA elbow in structure, recognition and evolution. *Life (Basel)* **6**, pii: E3.

4 Morales AJ, Swairjo MA and Schimmel P (1999) Structure-specific tRNA-binding protein from the extreme thermophile Aquifex aeolicus. *EMBO J* **18**, 3475–3483.

5 Cassio D and Waller JP (1971) Modification of methionyl-tRNA synthetase by proteolytic cleavage and properties of the trypsin-modified enzyme. *Eur J Biochem* **20**, 283–300.

6 Kalogerakos T, Dessen P, Fayat G and Blanquet S (1980) Proteolytic cleavage of methionyl transfer ribonucleic acid synthetase from *Bacillus stearothermophilus*: effects on activity and structure. *Biochemistry* **19**, 3712–3723.

7 Kohda D, Yokoyama S and Miyazawa T (1987) Functions of isolated domains of methionyl-tRNA synthetase from an extreme thermophile, *Thermus thermophilus* HB8. *J Biol Chem* **262**, 558–563.

8 Goldgur Y, Mosyak L, Reshetnikova L, Ankilova V, Lavrik O, Khodyreva S and Safro M (1997) The crystal structure of phenylalanyl-tRNA synthetase from thermus thermophilus complexed with cognate tRNAPhe. *Structure* **5**, 59–68.

9 Swairjo MA, Morales AJ, Wang CC, Ortiz AR and Schimmel P (2000) Crystal structure of trbp111: a structure-specific tRNA-binding protein. *EMBO J* **19**, 6287–6298.

10 Murzin AG (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J* **12**, 861–867.

11 Nomanbhoy T, Morales AJ, Abraham AT, Vörtler S, Giegé R and Schimmel P (2001) Simultaneous binding of two proteins to opposite sides of a single transfer RNA. *Nat Struct Biol* **8**, 344–348.

12 Renault L, Kerjan P, Pascalato S, Ménétrey J, Robinson J-C, Kawaguchi S, Vassylyev DG, Yokoyama S, Mirande M and Cherfils J (2001) Structure of the EMAPII domain of human aminoacyl-tRNA synthetase complex reveals evolutionary dimer mimicry. *EMBO J* **20**, 570–578.

13 Arnez JG, Augustine JG, Moras D and Francklyn CS (1997) The first step of aminoacylation at the atomic level in histidyl-tRNA synthetase. *Proc Natl Acad Sci USA* **94**, 7144–7149.

14 Giegé R, Sissler M and Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* **26**, 5017–5035.

15 Negrutskii BS and Deutscher MP (1991) Channeling of aminoacyl-tRNA for protein synthesis in vivo. *Proc Natl Acad Sci USA* **88**, 4991–4995.

16 Cho HY, Maeng SJ, Cho HJ, Choi YS, Chung JM, Lee S, Kim HK, Kim JH, Eom CY, Kim YG *et al.* (2015) Assembly of multi-tRNA synthetase complex via heterotetrameric glutathione transferase-homology domains. *J Biol Chem* **290**, 29313–29328.

17 Karanasios E and Simos G (2010) Building arks for tRNA: structure and function of the Arc1p family of non-catalytic tRNA-binding proteins. *FEBS Lett* **584**, 3842–3849.

18 Mirande M, Le Corre D and Waller JP (1985) A complex from cultured Chinese hamster ovary cells containing nine aminoacyl-tRNA synthetases. Thermolabile leucyl-tRNA synthetase from the tsH1 mutant cell line is an integral component of this complex. *Eur J Biochem* **147**, 281–289.

19 Havrylenko S and Mirande M (2015) Aminoacyl-tRNA synthetase complexes in evolution. *Int J Mol Sci* **16**, 6571–6594.

20 Quevillon S, Agou F, Robinson JC and Mirande M (1997) The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial monocyte-activating polypeptide II cytokine. *J Biol Chem* **272**, 32573–32579.

21 Kaminska M, Havrylenko S, Decottignies P, Gillet S, Le Maréchal P, Negrutskii B and Mirande M (2009) Dissection of the structural organization of the aminoacyl-tRNA synthetase complex. *J Biol Chem* **284**, 6053–6060.

22 Guigou L, Shalak V and Mirande M (2004) The tRNA-interacting factor p43 associates with mammalian arginyl-tRNA synthetase but does not modify its tRNA aminoacylation properties. *Biochemistry* **43**, 4592–4600.

23 Behrensdorf HA, van de Craen M, Knies UE, Vandenabeele P and Clauss M (2000) The endothelial monocyte-activating polypeptide II (EMAP II) is a substrate for caspase-7. *FEBS Lett* **466**, 143–147.

24 Shalak V, Kaminska M, Mitnacht-Kraus R, Vandenabeele P, Clauss M and Mirande M (2001) The EMAPII cytokine is released from the mammalian multisynthetase complex after cleavage of its p43/proEMAPII component. *J Biol Chem* **276**, 23769–23776.

25 Park SG, Choi EC and Kim S (2010) Aminoacyl-tRNA synthetase-interacting multifunctional proteins (AIMPs): a triad for cellular homeostasis. *IUBMB Life* **62**, 296–302.

26 Liang D, Halpert MM, Konduri V and Decker WK (2015) Stepping out of the cytosol: AIMp1/p43 potentiates the link between innate and adaptive immunity. *Int Rev Immunol* **34**, 367–381.

27 Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Cade S, Doherty R, Fischer S, Gajria B, Gao X, Gingle A *et al*. (2013) EuPathDB: the eukaryotic pathogen database. *Nucleic Acids Res* **41**(Database issue), D684–D691.

28 Cestari I, Kalidas S, Monnerat S, Anupama A, Phillips MA and Stuart K (2013) A multiple aminoacyl-tRNA synthetase complex that enhances tRNA-aminoacylation in African trypanosomes. *Mol Cell Biol* **33**, 4872–4888.

29 Simos G, Sauer A, Fasiolo F and Hurt EC (1998) A conserved domain within Arc1p delivers tRNA to aminoacyl-tRNA synthetases. *Mol Cell* **1**, 235–242.

30 Karanasios E, Simader H, Panayotou G, Suck D and Simos G (2007) Molecular determinants of the yeast Arc1p-aminoacyl-tRNA synthetase complex assembly. *J Mol Biol* **374**, 1077–1090.

31 Deinert K, Fasiolo F, Hurt EC and Simos G (2001) Arc1p organizes the yeast aminoacyl-tRNA synthetase complex and stabilizes its interaction with the cognate tRNAs. *J Biol Chem* **276**, 6000–6008.

32 Graindorge JS, Senger B, Tritch D, Simos G and Fasiolo F (2005) Role of Arc1p in the modulation of yeast glutamyl-tRNA synthetase activity. *Biochemistry* **44**, 1344–1352.

33 Simos G, Segref A, Fasiolo F, Hellmuth K, Shevchenko A, Mann M and Hurt EC (1996) The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl- and glutamyl-tRNA synthetases. *EMBO J* **15**, 5437–5448.

34 Galani K, Grosshans H, Deinert K, Hurt EC and Simos G (2001) The intracellular location of two aminoacyl-tRNA synthetases depends on complex formation with Arc1p. *EMBO J* **20**, 6889–6898.

35 Frechin M, Enkler L, Tetaud E, Laporte D, Senger B, Blancard C, Hammann P, Bader G, Clauder-Münster S, Steinmetz LM *et al*. (2014) Expression of nuclear and mitochondrial genes encoding ATP synthase is synchronized by disassembly of a multisynthetase complex. *Mol Cell* **56**, 763–776.

36 van Rooyen J, Murat J-B, Hammoudi P-M, Kieffer-Jaquinod S, Coute Y, Sharma A, Pelloux H, Belrhali H and Hakimi M-A (2014) Assembly of the novel five-component apicomplexan multi-aminoacyl-tRNA synthetase complex is driven by the hybrid scaffold protein Tg-p43. *PLoS One* **9**, e89487.

37 Bour T, Mahmoudi N, Kapps D, Thiberge S, Bargieri D, Ménard R and Frugier M (2016) Apicomplexa-specific tRip facilitates import of exogenous tRNAs into malaria parasites. *Proc Natl Acad Sci USA* **113**, 4717–4722.

38 Castro de Moura M, Miro F, Han JM, Kim S, Celada A and Ribas de ouplana L (2011) Entamoeba lysyl-tRNA synthetase contains a cytokine-like domain with chemokine activity towards human endothelial cells. *PLoS Negl Trop Dis* **5**, e1398.

39 Gilchrist CA and Petri WA Jr (2009) Using differential gene expression to study *Entamoeba histolytica* pathogenesis. *Trends Parasitol* **25**, 124–131.

40 Kaminska M, Deniziak M, Kerjan P, Barciszewski J and Mirande M (2000) A recurrent general RNA binding domain appended to plant methionyl-tRNA synthetase acts as a cis-acting cofactor for aminoacylation. *EMBO J* **19**, 6908–6917.

41 Havrylenko S, Legouis R, Negrutskii B and Mirande M (2010) Methionyl-tRNA synthetase from *Caenorhabditis elegans*: a specific multidomain organization for convergent functional evolution. *Protein Sci* **19**, 2475–2484.

42 Wakasugi K and Schimmel P (1999) Two distinct cytokines released from a human aminoacyl-tRNA synthetase. *Science* **284**, 147–151.

43 van Horssen R, Eggermont AM and ten Hagen TL (2006) Endothelial monocyte-activating polypeptide-II

and its functions in (patho)physiological processes. *Cytokine Growth Factor Rev* **17**, 339–348.
44 Wakasugi K and Schimmel P (1999) Highly differentiated motifs responsible for two cytokine activities of a split human tRNA synthetase. *J Biol Chem* **274**, 23155–23159.
45 Siddiqui R and Khan NA (2012) Biology and pathogenesis of Acanthamoeba. *Parasit Vectors* **5**, 6.
46 Wakefield WS, Powell MJ, Letcher PM, Barr DJ, Churchill PF, Longcore JE and Chen SF (2010) A molecular phylogenetic evaluation of the spizellomycetales. *Mycologia* **102**, 596–604.
47 Berger L, Speare R, Daszak P, Green DE, Cunningham AA, Goggin CL, Slocombe R, Ragan MA, Hyatt AD, McDonald KR *et al.* (1998) Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proc Natl Acad Sci USA* **95**, 9031–9036.
48 Yang X, Matern HT and Gallwitz D (1998) Specific binding to a novel and essential Golgi membrane protein (Yip1p) functionally links the transport GTPases Ypt1p and Ypt31p. *EMBO J* **17**, 4954–4963.
49 Frantz JD and Gilbert W (1995) A novel yeast gene product, G4p1, with a specific affinity for quadruplex nucleic acids. *J Biol Chem* **270**, 20692–20697.
50 Rhodes D and Lipps HJ (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* **43**, 8627–8637.
51 Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, Rees B, Thierry JC and Moras D (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* **252**, 1682–1689.
52 Giessen TW, Altegoer F, Nebel AJ, Steinbach RM, Bange G and Marahiel MA (2015) A synthetic adenylation-domain-based tRNA-aminoacylation catalyst. *Angew Chem Int Ed Engl* **54**, 2492–2496.
53 Ruggiero MA, Gordon DP, Orrell TM, Bailly N, Bourgoin T, Brusca RC, Cavalier-Smith T, Guiry MD and Kirk PM (2015) A higher level classification of all living organisms. *PLoS One* **10**, e0119248.
54 Kao J, Fan YG, Haehnel I, Brett J, Greenberg S, Clauss M, Kayton M, Houck K, Kisiel W, Seljelid R *et al.* (1994) A peptide derived from the amino terminus of endothelial-monocyte-activating polypeptide II modulates mononuclear and polymorphonuclear leukocyte functions, defines an apparently novel cellular interaction site, and induces an acute inflammatory response. *J Biol Chem* **269**, 9774–9782.
55 Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

## V. OBJECTIFS DE MA THESE

Bour, *et al.*, 2016, met en évidence que les sporozoïtes importent des ARNt exogènes, cependant c'est au stade sanguin que le développement du parasite tRip-KO est réduit. Dans ce contexte, mon projet de thèse consiste à (i) identifier les ARNt qui sont les plus susceptibles d'être importés (*in vitro*) et (ii) démontrer la présence des ARNt de l'hôte dans le parasite sauvage au stade sanguin (*in vivo*).

### 1. La sélectivité de tRip *in vitro*

La protéine tRip se lie à l'ensemble des ARNt humains avec une affinité élevée et apparemment sans aucune spécificité de séquence (Bour, *et al.*, 2016). Je suis partie aux Etats Unis pour utiliser la technique MIST (Microarray Identification of Shifted tRNAs) qui a été conçue par R. Geslain. Cette technique, qui combine des expériences de retard sur gel (sélection) avec des analyses sur puces à ARNt pour étudier la formation des complexes ARN/protéine, représente la meilleure approche pour identifier un biais potentiel dans la formation des complexes entre tRip et les ARNt de l'hôte. Mes résultats montrent clairement que tRip a plus d'affinité pour certains ARNt et que cette sélection est due à la présence des bases modifiées. L'interaction préférentielle de tRip avec certains ARNt, suggère qu'ils pourraient être importés préférentiellement et nous donnent des indications précieuses sur le rôle des ARNt transportés à l'intérieur du parasite. En parallèle, j'ai utilisé la technique du SELEX pour identifier les motifs nucléotidiques capables de se lier à tRip avec une affinité comparable aux ARNt.

### 2. Présence des ARNt de l'hôte dans le parasite au stade sanguin

La source d'ARNt de l'hôte dans la stade sanguin est énigmatique puisque les érythrocytes matures ne contiennent pas ou peu d'ARNt. En effet, les globules rouges matures ne possèdent plus de noyau et ne sont plus actifs en synthèse protéique. De plus, la présence d'une vacuole parasitophore isole le parasite du cytosol de l'hôte et n'est donc pas propice à l'import des ARNt de l'hôte dans le parasite. J'ai initié des expériences pour rechercher la

présence d'ARNt de l'hôte dans les parasites sauvages au cours du stade sanguin. Ce travail demande une grande dextérité pour purifier les parasites en évitant toute contamination par des ARNt de l'hôte non importés. Les premiers résultats sont très encourageants mais des traces d'ARNt de l'hôte dans les échantillons du parasite tRip-KO (contrôle négatif) indiquent que le protocole de purification des parasites nécessite encore quelques améliorations.

## 3. La tyrosylation dans l'apicoplaste de *Plasmodium falciparum*

En parallèle de mon projet principal, j'ai étudié la Tyrosyl-ARNt synthétase apicoplastique, l'enzyme qui catalyse spécifiquement l'attachement de la tyrosine sur l'ARNt$^{Tyr}$ dans l'apicoplaste du parasite. Le gène de la TyrRS apicoplastique a été cloné et exprimé dans *E. coli*. Cette stratégie a permis de produire suffisamment de protéine recombinante soluble et active pour étudier ses caractéristiques enzymatiques *in vitro*. L'ARNt$^{Tyr}$ apicoplastique, a été produit par transcription *in vitro* et j'ai vérifié sa structure par cartographie aux sondes chimiques et enzymatiques. J'ai également produit plusieurs variantes de cet ARNt et de la TyrRS qui m'ont permis de révéler un mode d'interaction inédit entre les deux partenaires. Mes résultats ont permis de mettre en évidence un jeu d'identité tyrosine dans l'apicoplaste de *Plasmodium* unique qui complète les données nombreuses sur l'évolution de ce système d'aminoacylation.

# RESULTATS

## I. IDENTIFICATION DES ELEMENTS D'IDENTITE DE L'ARNT$^{TYR}$ APICOPLASTIQUE

Les éléments d'identité sont normalement conservés tout au long de l'évolution. Cependant, le système de tyrosylation (ARNt$^{Tyr}$/TyrRS) est une exception. En effet, il existe une barrière d'espèce qui empêche la tyrosilation croisée entre les bactéries, les archaebactéries et les eucaryotes. Elle est constituée de la première paire de bases située à l'extrémité du bras accepteur de l'ARNt$^{Tyr}$, $G_1C_{72}$ chez les bactéries et $C_1G_{72}$ chez les archaebactéries et les eucaryotes (Revu dans Bonnefond, *et al.*, 2005). De manière surprenante, l'ARNt$^{Tyr}$ de l'apicoplaste de *Plasmodium* possède une paire de bases unique et conservée chez tous les *Plasmodium* : $A_1U_{72}$. Est-ce que cette paire de base fait partie du jeu d'élément d'identité du système tyrosine de l'apicoplaste ? Est-ce que comme dans le cas de la tyrosylation dans les mitochondries de vertébrés, cette paire de base ne joue aucun rôle dans la reconnaissance spécifique de l'ARNt$^{Tyr}$ ? Est-ce que d'autres éléments d'identité sont impliqués ?

De plus, la Tyrosyl-ARNt synthétase de l'apicoplaste contient plusieurs insertions qui rendent l'expression d'une protéine recombinante particulièrement difficile. La première étape de cette étude a donc consisté à trouver des conditions d'expression favorables.

### *Plasmodium* apicoplast tyrosyl-tRNA synthetase recognizes an unusual simplified identity set in cognate tRNA$^{Tyr}$

Marta Cela[1], Caroline Paulus[1], Manuel A. S. Santos[2], Gabriela Moura[2], Magali Frugier[1]* and Joëlle Rudinger-Thirion[1]

[1] UPR 9002 Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, F-67084 Strasbourg Cedex, France

[2] Department of Medical Sciences and Institute of Biomedicine - iBiMED, University of Aveiro, 3810-193 Aveiro, Portugal

*To whom correspondence should be addressed

Tel: 33 3 88 41 71 02

Fax: 33 3 88 60 22 18

Email: m.frugier@ibmc-cnrs.unistra.fr

**ABSTRACT**

The life cycle of *Plasmodium falciparum*, the agent responsible for malaria, depends on both cytosolic and apicoplast translation fidelity. Apicoplast aminoacyl-tRNA synthetases (aaRS) are bacterial-like enzymes devoted to the organellar tRNA aminoacylation. They are all encoded by the nuclear genome and reach their compartment only after their cytosolic biosynthesis. Apicoplast aaRSs are characterized by a high content of idiosyncratic insertions, which understanding remains elusive because expression of soluble recombinant proteins containing these insertions is challenging. Moreover, the A/T rich content of the *Plasmodium* genome leads to A/U rich apicoplast tRNA substrates. Here we focus on *P. falciparum* apicoplast tyrosyl-tRNA synthetase (*Pf*-apiTyrRS) and its cognate *Pf*-apitRNA$^{Tyr}$ substrate that displays structural plasticity. Cloning and expression strategies used to obtain an active and functional *Pf*-apiTyrRS *in vitro* are reported. When compared to other tyrosine systems from different *Phylae*, *Pf*-apiTyrRS behaves differently. Functional analyses established that only three weak identity elements in the apicoplast tRNA$^{Tyr}$ promote specific recognition by the *Pf*-apiTyrRS and that phylogenetically conserved tyrosine identity elements usually found in the tRNA acceptor stem are excluded from this set. This finding brings to light an unusual behavior of a tyrosylation system and suggests that that *Pf*-apiTyrRS uses primarily negative recognition elements to direct tyrosylation specificity.

## INTRODUCTION

*Plasmodium falciparum* is the agent of the most severe form of malaria in humans. Rapid constitutive growth and expansion of the parasite is highly dependent on the continuous synthesis of proteins in its cytosolic and organellar compartments. Indeed, *Plasmodium* contains three genomes: nuclear, apicoplast (a relic chloroplast) and mitochondrial. These genomes require dedicated translation mechanisms to function (1), even if translation is not yet explicitly demonstrated in the mitochondria. The *Plasmodium* apicoplast is essential and is involved in the synthesis of fatty acids, isoprenoid precursors and heme (2, 3). It has a 35 kb circular genome encoding 30 protein genes such as RNA polymerase subunits, elongation factor Tu, several ribosomal proteins as well as ribosomal RNAs and transfer RNAs (tRNAs), all involved in apicoplast transcription and translation (4). However, most of the proteins essential for the apicoplast functions are encoded by the nuclear genome and are imported into the organelle after translation. The *P. falciparum* genome contains more than 450 protein genes with a targeting signal for the apicoplast (5), among them are the apicoplast aaRS.

Generally, there are 20 different aaRSs in each protein translation compartment, one for each of the twenty amino acids (6). In *Plasmodium*, 37 nuclear genes encode 36 aaRSs (the cytosolic Phenylalanyl-tRNA synthetase is a heterodimer) (1, 7); 17 are exclusively cytosolic, 20 possess an apicoplast bi-partite targeting sequence among which, 4 are dual-targeted to the apicoplast and the cytosol (alanyl-, cysteinyl-, glycyl- and threonyl-tRNA synthetases) and one is putatively mitochondrial (8).

Although aaRSs perform the same reaction, they are structurally different. Depending on the architecture of the active site and the tRNA binding mode, the aaRS are divided into two structural classes, with 10 enzymes in each class (6 , 9 , 10). In addition to the catalytic domain, they typically present a tRNA-binding domain, often recognizing the anticodon triplet of the tRNA. This organization holds true for organellar aaRSs with however some differences in size, oligomerization and at the functional levels (11). Apicoplast aaRSs are evolutionarily conserved enzymes, yet previous studies have highlighted the presence of numerous uncharacterized insertions, which therefore increase significantly their size (7). The presence of these insertions could explain why apicoplast aaRSs are poorly studied despite their interest in the development of new anti-malarial drugs (12 , 13 , 14)

It is accepted that the fidelity of protein synthesis depends mostly on the formation of correct aminoacyl-tRNAs by aaRSs (6 , 15). Thus, a given aaRS discriminates one amino acid among the 20 canonical amino acids, as well as its homologous tRNAs. The specificity of this

process is mainly triggered by identity elements (nucleotides or structural motifs) scattered in the tRNA structure. As positive identity elements, these motifs are specific for each tRNA/aaRS pair. Moreover, negative identity elements play also an essential role by preventing the binding of noncognate aaRSs to a given tRNA. tRNA identity sets have been characterized for most eubacterial, but fewer archaeal, eukaryal and organellar aminoacylation systems (16 , 17). In view of the conserved canonical structure of tRNAs, the nature of identity elements are generally preserved during evolution, however, exceptions have been reported. In particular, the tyrosine system is characterized by a robust phylogenetic barrier that prevents cross-reactivity between partners of different species (e.g.18 , 19 , 20).

Guided by the above conceptual background, we characterized the structural features required to specifically aminoacylate the $tRNA^{Tyr}$ encoded by the apicoplast genome (*Pf*-$apitRNA^{Tyr}$) by the apicoplast tyrosyl-tRNA synthetase of *Plasmodium falciparum* (*Pf*-apiTyrRS). It first required the engineering of the *Pf*-apiTyrRS gene to produce a soluble recombinant protein containing two *Plasmodium*-specific insertions (Figure S1). Then, a systematic study of nucleotide replacement and a comparative examination of the activity of tRNA variants established the rules that govern the recognition of *Pf*-$apitRNA^{Tyr}$ by its cognate synthetase. Our analysis demonstrates that the identity elements of the tyrosine system are reduced in number and strength, in *Pf*-$apitRNA^{Tyr}$ and that again they are not conserved evolutionarily.

## MATERIAL AND METHODS

### *Cloning and purification of P. falciparum apicoplast TyrRS*

The genomic sequence of *Pf*-apiTyrRS was retrieved from PlasmoDB (21) by sequence homology with the human mitochondrial TyrRS (EAW88518.1, *Hs*-mitoTyrRS) and *Thermus thermophilus* TyrRS (AEG33811.1) (3). The gene (PF3D7_1117500) codes for a 561 amino acids long protein. The *Pf*-apiTyrRS gene was amplified by PCR from a *P. falciparum* cDNA library (provided by Dr. H. Vial, Montpellier), sequenced, and cloned into pQE30 (Qiagen) to produce a protein with a 6-histidine fusion tag at its N-terminus. The *Pf*-apiTyrRS recombinant protein is deprived of its N-terminal apicoplast targeting signal and covers amino acids 25 to 561.

Besides the *P. falciparum* nucleotide sequence of *Pf*-apiTyrRS (Endo), two other nucleotide sequences, both encoding the same amino acid sequence were designed (Figure S2). We produced an "optimized" version (Opt) of the gene based on the *Escherichia coli* codon usage (designed by Genscript) and a "harmonized" version (Harm) of the gene using the ANACONDA software (22 , 23). Both genes were synthesized by GenScript (https://www.genscript.com) and cloned into the pQE30 plasmid.

Furthermore, a truncated variant of Harm *Pf*-apiTyrRS gene lacking the C-terminal S4-like domain (residues 460 to 561) was cloned in pQE70 with a C-terminal 6-His tag.

Overexpression of all recombinant proteins was performed at 18°C O/N in LB medium containing 1 mM IPTG, and the purification of 6-His tagged proteins was done on Ni-NTA resin according to the manufacturer's instructions (Qiagen). Purified enzymes were dialyzed against 25 mM HEPES-KOH pH 7.5, 25 mM KCl, 50% glycerol and kept at -20°C until use. Proteins were quantified, and their enzymatic activities were measured by *in vitro* aminoacylation of native *E. coli* $tRNA^{Tyr}$.

### *Sequence analysis*

Sequence alignments were computed with Tcoffee software (24), and CLUSTALW (25). The prediction of *Pf*-apiTyrRS secondary structure was achieved with the PredictProtein software (26). The PlasmoAP algorithm (21) was used to confirm the presence of an apicoplast targeting signals and predict the cleavage site. The *E. coli* codon usage database was from (https://openwetware.org/mediawiki/index.php?title=Escherichia_coli/Codon_usage&oldid=820473).

***Preparation of tRNA molecules***

Native *E. coli* $tRNA^{Tyr}$ (Genbank CP010229.1) was purchased from Sigma-Aldrich. Transcripts corresponding to wild-type (PF3D7_API00700) and mutated *Pf*-$apitRNA^{Tyr}$ were synthesized by *in vitro* transcription using the «transzyme» method (27). Briefly, the synthetic genes contain the T7 RNA polymerase promoter followed by a hammerhead ribozyme and the tRNA gene sequences. They were cloned in pUC119, linearized with BstNI (to produce a 3' CCA sequence after transcription) and transcribed *in vitro* by the T7 RNA polymerase. The auto-catalytically cleaved transcripts were purified on 12% preparative denaturating polyacrylamide gels. Similarly, we produce transcripts corresponding to the wild-type *P. falciparum* apicoplast $tRNA^{Ser}_{GCU}$ (PF3D7_API00800).

***Procedures for structural analysis of free and TyrRS-complexed tRNAs***

Lead cleavage (10 µl) was performed as in (28) : 5'-labeled *Pf*-$apitRNA^{Tyr}$ wild-type transcript (80000 cpm, 1 µM) was incubated in 50 mM Tris-acetate, pH 7.5, 5 mM magnesium acetate, 50 mM potassium acetate. A freshly prepared $Pb(OAc)_2$ solution in $H_2O$ was added to reach final concentrations of 1, 5 and 10 mM. Samples were incubated for 6 min at 25°C; the reactions were stopped in ice in the presence of 5 µl EDTA 0.1 M and 1 µg glycogen. RNA fragments were ethanol precipitated; pellets were washed twice with ethanol 70%, vacuum-dried, dissolved in the loading mix (90% formamide, 0.5% EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue) and loaded on a 12% denaturing gel.

Enzymatic probing (10 µl) was performed as in (28). Incubations contained 80000 cpm of 5'-labeled *Pf*-$apitRNA^{Tyr}$ transcript supplemented with 1 µM cold transcript in 20 mM $MgCl_2$, 100 mM NaCl, 50 mM Hepes-NaOH pH 7.5 for 10 min at 25°C and in the presence of T1 (0.2 U), S1 (5.4 U) and V1 ($7x10^{-4}$ U) nucleases. For S1 probing, 1 mM $ZnCl_2$ was added. A control experiment without nuclease was performed in parallel. Reactions were stopped by addition of 20 µl of Stop Mix (0.6 M NaOAc, 4 mM EDTA, 0.1 mg/ml total tRNA and 1 µg glycogen). After ethanol precipitation, the pellets were treated as described for lead probing. In parallel, T1 hydrolysis and alkaline degradation were performed on the *Pf*-$apitRNA^{Tyr}$ transcript in denaturing conditions to assign the bands accurately.

Footprinting assays (10 µl) were performed in the same conditions as above in the absence or presence of 5.7 µM of *Pf*-apiTyrRS. The complex tRNA/TyrRS was incubated for 6 min at 25°C, before 0.2 U T1 or 5.4 U S1 (supplemented with 1 mM $ZnCl_2$ for S1 cleavage) were added. Incubation was extended for 8 min at 25°C and the reaction stopped by phenol extraction.

***Aminoacylation assays***

Tyrosylation assays were performed at 37°C in 50 mM Hepes-KOH (pH 7.6), 25 mM KCl, 12 mM $MgCl_2$, 2.5 mM ATP, 0.2 mg/ml bovine serum albumin, 1 mM spermine, 10 µM L-[$^{14}$C]-tyrosine at 486 mCi/mmol (Perkin Elmer) and appropriate amounts of either native or transcribed $tRNA^{Tyr}$ molecules. *Pf*-apiTyrRS (50 to 100 nM, diluted in 100 mM Hepes-NaOH pH 7.4, 1 mM DTT, 5 mg/ml BSA and 10% glycerol) was added to start the reaction (29). Apparent kinetic parameters were determined from Lineweaver-Burk plots.

**RESULTS**

***Identification of P. falciparum apicoplast* TyrRS *gene***

We identified a gene coding for a 561 amino acid *Plasmodium* protein (PF3D7_1117500) corresponding to *Pf*-apiTyrRS containing an N-terminal catalytic domain and an anticodon binding domain at its C-terminus (Figure 1A). The catalytic domain (amino acids 1-328, Rossman-fold) presents both class I aaRS specific motifs (HNGL and KYSKS). As expected, the C-terminal domain of *Pf*-apiTyrRS presents the typical α-helical domain and the S4-like region that are both specific features found only in bacterial and mitochondrial TyrRSs (reviewed in (19)).

The PlasmoAP algorithm that predicts apicoplast targeting signals indicates that the N-terminal extremity of the protein contains "very likely" an apicoplast targeting signal. In addition, predictions of secondary structures indicated that the 21$^{st}$ residues of the protein encompassed a helix (Figure S1). This knowledge led to the deletion of the first 24 amino acids from the N-terminus in the recombinant *Pf*-apiTyrRS.

*Pf*-apiTyrRS is longer than its homologs (Figure 1A). This is the consequence of 2 insertions of 19 and 56 amino acids found in the catalytic and the anticodon-binding domains, respectively. Sequence alignments with several other *Plasmodium* apicoplast TyrRSs revealed conserved localizations of these insertions whereas the sizes and sequences vary significantly (Figure S1). The insertion located in the anticodon-binding domain is characterized by single amino acid repeats (30).

***Production of a functional recombinant P. falciparum apicoplast TyrRS***

The *Pf*-apiTyrRS could be expressed in *E. coli* directly from the *P. falciparum* endogenous gene (Endo) but led to a reduced yield of soluble protein after affinity purification (500 ml of culture yielded less than 1 mg of protein).

An alternative strategy was used to improve the production of soluble *Pf*-apiTyrRS in the *E. coli* heterologous expression system. Two different synthetic gene sequences encoding the same amino acid chain were cloned (Figure S2). On the one hand, the DNA sequence was changed using standard codon optimization rules (GenScript, Opt). On the other hand, the DNA sequence was "harmonized" (Harm) using the bioinformatics application for gene primary structure analysis ANACONDA. This program uses statistical methods to analyze not only the codon usage but also the codon context (degree of association, contexts and clustering) on a genomic scale (22 , 23). In other words, it takes into account the rules governing the evolution of the codon bias in *P. falciparum* to design a new nucleotide sequence adapted to the *E. coli* translational apparatus. After harmonization, the main differences appeared at the level of leucine codons which were all replaced by the CUA, the rarest amongst the six leucine codon ($3.85^{0}/_{00}$)

Gene optimization (Opt) did not change the protein expression significantly compared to the *Plasmodium* endogenous DNA sequence (Endo), however, it reduced the solubility of *Pf*-apiTyrRS since the Western Blot signal corresponding to the protein is decreased in the corresponding soluble extract (Figure 1B). This was verified by the poor protein yield obtained after purification (Figure 1C). The best expression yield was obtained with the Harm gene, with 3 times more purified recombinant *Pf*-apiTyrRS than with the Endo gene (Figure 1C). Moreover, comparative aminoacylation assays (Figure 1D) using native *E. coli* $tRNA^{Tyr}$ demonstrated that the produced enzymes were not functionally equivalent. The Harm *Pf*-apiTyrRS is significantly more efficient in aminoacylation than the 2 other preparations. This indicates that gene harmonization not only increased the solubility and thus the purification yield of *Pf*-apiTyrRS, but also improved the correct folding of the recombinant protein.

*Pf*-apiTyrRS displays an S4-like domain at its C-terminus, specific for bacterial and mitochondrial TyrRSs (Figure 1A). In general, the elimination of this domain leads to more soluble TyrRSs although characterized by a lower affinity for $tRNA^{Tyr}$ (e.g. (32)), since the recognition of the tRNA variable region is disrupted (33). In contrast to these prokaryotic-type TyrRSs, the truncation of the S4-like domain of *Pf*-apiTyrRS led to an inactive enzyme unable of tRNA charging because it does not catalyze tyrosine activation in the presence of ATP (the first step of the aminoacylation reaction) (Figure S3). It suggests that the deletion of the C-terminal domain affects the folding of the catalytic site.

Thus, *Pf*-apiTyrRS (25-561) expressed from the harmonized construct was used to determine the aminoacylation efficiencies of the different mutants of *Pf*-$apitRNA^{Tyr}$.

***Sequence peculiarities in P. falciparum apicoplast tRNA$^{Tyr}$***

The *Pf*-apiTyrRS natural substrate is the unique tRNA$^{Tyr}$ encoded by the apicoplast genome (Figure 2A-B). *Pf*-apitRNA$^{Tyr}$ is composed of 70% A/U residues (in both stems and loops) reflecting the rich A/T composition of the *P. falciparum* genome (4). *Pf*-apitRNA$^{Tyr}$ displays (i) the phylogenetically conserved A73 discriminator base, (ii) an α3β4 D-loop with a non-canonical G13-A22 base pair in the D-stem and (iii) like the eubacterial tRNA$^{Tyr}$, it belongs to class 2 tRNAs, defined as tRNAs with a large variable region (Figure 2B). Interestingly, *Pf*-tRNA$^{Tyr}$ is characterized by the presence of A1-U72 base pair at the top of the acceptor stem. This characteristic is only conserved in *Plasmodium* apicoplast tRNA$^{Tyr}$ (Figure 2A), while tRNA$^{Tyr}$ from bacteria, mitochondria and even chloroplasts are characterized by a G1-C72 base pair and archaeal and eukaryotic tRNA$^{Tyr}$ contain a C1-G72 base pair (Figure 2C).

***Cloverleaf folding of P. falciparum apicoplast tRNA$^{Tyr}$***

The 87 nucleotide-long *Pf*-apitRNA$^{Tyr}$ was produced as a transcript deprived of modified bases. Probing experiments were performed with PbOAc and nucleases (Figure 3A) to verify that the high proportion of A and U residues and the absence of posttranscriptional modifications do not hinder the formation of the canonical cloverleaf fold. Nucleases S1 and T1, as well as PbOAc, are specific for single-stranded regions whereas nuclease V1 is specific for double-stranded and highly structured regions. The similarities between T1 patterns in both native (T1) and denaturing (G) conditions strongly suggest that *Pf*-tRNA$^{Tyr}$ transcript is flexible (Figure 3A). This is confirmed by the lead cleavages, which occur throughout the sequence, though with the strongest cuts in the loops (D-, anticodon, T- loops and the variable region). Nevertheless, as expected for a cloverleaf fold, strongest S1 and V1 cuts are found only in the anticodon loop and the anticodon-stem, respectively, indicating that the transcript is indeed folded. Moreover, the moderate accessibilities of $A_{57}$ and $U_{20A}$ towards V1 confirm the presence of the T/D loop tertiary interaction, whereas conflicting V1 and S1 cuts in the variable region confirm a fluctuating structure in this domain. Altogether, these probing data are in agreement with a cloverleaf fold of the molecule and indicate an intrinsic structural plasticity, reminiscent of what was observed for 'light' mitochondrial tRNAs (28, 42). Yet, the correct folding of *Pf*-apitRNA$^{Tyr}$ was confirmed since this transcript is aminoacylated by the homologous *Pf*-apiTyrRS with the same catalytic efficiency than the native *E. coli* tRNA$^{Tyr}$ (Table 1).

***Extensive binding proximities of P. falciparum apicoplast tRNA$^{Tyr}$ with its cognate TyrRS***

Nucleases S1 and V1 were also used in footprinting experiments (Figure 3B) to detect the regions of *Pf*-apitRNA$^{Tyr}$ that are protected in the presence of *Pf*-apiTyrRS. Both the anticodon loop and the variable domain of the *Pf*-apitRNA$^{Tyr}$ transcript are strongly protected from RNases' cleavages and mirror what was observed in the crystallographic structure of the TyrRS/tRNATyr complex from *T. thermophilus* (33). However, the protections detected in the D-domain of *Pf*-apitRNA$^{Tyr}$ did not corroborate the bacterial recognition pattern (33). The disappearance of accessibilities in the D-loop is the consequence either of direct contact with the synthetase, or of an indirect effect due to a conformational change in the tRNA when it interacts with the enzyme.

***Looking for identity elements in P. falciparum apicoplast tRNA$^{Tyr}$***

TyrRSs display species-specific tRNA recognition (35, 36, 37, 41, 43) (summarized in Figure 2C with the references therein). This discrimination is universally insured by the tRNA$^{Tyr}$ discriminator base A73 and the anticodon nucleotide G34. Anticodon nucleotides U35 and A36 contribute also to tyrosylation identity, yet with variable strengths in eukarya, eubacteria, and mitochondria and are even marginal in archebacteria tyrosine identity. Notably, the identity base-pair, G1–C72, located at the top of the acceptor stem of eubacteria tRNA$^{Tyr}$ is replaced by the identity base pair C1–G72 in archeabacteria and eubacteria tRNA$^{Tyr}$ and the large variable region which is unique to eubacteria tRNA$^{Tyr}$ is essential to direct its tyrosylation.

Therefore, 18 *Pf*-apitRNA$^{Tyr}$ mutants were designed to test the extremity of the acceptor stem (discriminator base and the 1-72 base-pair), the anticodon triplet and the variable region for their involvement in apicoplast tyrosylation (Figure 4). Kinetic parameters for tyrosylation were determined and compared to those obtained for the wild-type molecule (Table 1).

*Acceptor stem:* Unexpectedly, replacing the discriminator A73 by G, C or U affects tyrosylation by only 2 to 4.2-fold, suggesting that these mutations do not affect the recognition of the tRNA by *Pf*-apiTyrRS but rather modulate to some extent the structure of the tRNA near the catalytic site. Indeed, it has been shown that a pyrimidine residue at position 73 might influence the stability of the acceptor stem extremity (44).

In contrast to the vast majority of tRNA$^{Tyr}$, where the first base pair in the acceptor stem is G1-C72 (in eubacteria) or C1-G72 (in archaebacteria and eukarya), all *Plasmodium* apicoplast tRNA$^{Tyr}$ contain a conserved A1-U72 (Figure 3A). A1-U72 was thus changed to

G-C and C-G, (Figure 4 and Table 1). *Pf*-apiTyrRS aminoacylates the wild-type transcript but also G1-C72 and C1-G72 variants with similar kinetic values (1, 0.8 and 1.8-fold). It indicates that the first base pair of the acceptor stem does not participate in tyrosine identity, providing the second example of a TyrRS lacking specificity toward the 1-72 base pair (19).

Because the recognition of the acceptor stem could be shifted, we looked for other residues that might play a role in tyrosine identity. Base pairs 2-71 and 3-70 were thus mutated (Figure 4 and Table 1). The recombinant *Pf*-apiTyrRS aminoacylated these variants with kinetic parameters similar to the wild-type transcript characterized by a loss in efficiency increased only 2.6-fold. Together, these data show the absence of tyrosine identity elements in the acceptor stem and suggest that *Pf*-apiTyrRS recognizes only the ribose-phosphate backbone in this region of the tRNA molecule.

*Anticodon triplet: R*esidues 34, 35 and 36 were individually mutated in *Pf*-apitRNA$^{Tyr}$. Only the variant where C34 replaced G34 showed a substantial loss in tyrosylation (36-fold). Mutations at position 35 affected the catalytic efficiency by 1 to 7.9-fold and mutations at position 36 had nearly no effect (2.8 to 3.6-fold) (Table 1). Changing the tyrosine anticodon into a serine GCU ($C_{35}U_{36}$) lead to an additive effect (7.9 x 2.8 = 22.12) and leads to a moderate 19-fold decrease in tyrosylation efficiency.

*Variable region and D-loop:* Shortening the variable region (DVr) decreased the catalytic efficiency by a factor of 6 as compared to wild-type *Pf*-apitRNA$^{Tyr}$ (Table 1) but replacing it with the variable region from *Pf*-apitRNA$^{Ser}$ (SerVr) had a weaker effect (2.7-fold). In *T. thermophilus* tRNA$^{Tyr}$, $A_{20B}$ in the D-loop base pairs with $U_{47i}$, resulting in a precise orientation of the variable region (30) for its optimal recognition by TyrRS. Since *Pf*-apitRNA$^{Tyr}$ sequence displays both $A_{20B}$ and $U_{47i}$, the same tertiary interaction could form (Figure 3A). The replacement of $A_{20B}$ by $U_{20B}$ in the *Pf*-apitRNA$^{Tyr}$ D-loop should thus eliminate this interaction, yet, this mutant showed no loss in tyrosylation (1.6-fold) (Table 1). Finally, both the D-loop and the variable region of *Pf*-apitRNA$^{Tyr}$ were changed for the *Pf*-apitRNA$^{Ser}$ (SerVr+D-l) sequences and did not lead to any effect in the tyrosylation efficiency (2.1-fold). Together these mutants demonstrate that in the apicoplast the presence of a long variable region plays a weak but significant role in tyrosylation identity, however neither the sequence nor the orientation of this long variable region is involved.

## DISCUSSION

### *Expression of Pf-apiTyrRS*

*Plasmodium* proteins are significantly longer than their homologs because they contain many insertions (4). Neither the synthesis nor the functions of these insertions are understood (30); the presence of long single amino acid repeats often hinders the solubility of the recombinant protein, but they cannot be removed under penalty of obtaining an inactive protein (*e.g*, (45)). It has been suggested that in *Plasmodium*, the translation of these additional sequences locally reduces the rate of ribosomes and is used to regulate protein co-translational folding (7). Their presence could explain why apicoplast aaRSs are poorly investigated despite their interest in the development of new anti-malarial drugs (14). Insertions are more frequent and more extended in apicoplast than in cytosolic *Plasmodium* aaRSs (7). The most extreme situation is that of the apicoplast leucyl-tRNA synthetase, that contains 9 insertions ranging from 30 to 121 amino acids. Up to date, only three apicoplast aaRSs from *P. falciparum* were cloned, expressed and characterized, namely lysyl- (LysRS) (46), glutamyl- (GluRS) (47) and the dual-targeted cysteinyl-tRNA synthetases (CysRS) (48). While both CysRS and GluRS do not contain additional sequences, LysRS is the only apicoplast insertion-containing aaRS that has been studied to date.

Translation is influenced by the choice of synonymous codons, which specify the same amino acid, but which differ in their decoding properties (49). Thus, the primary structure of the mRNA contains information that affects the efficiency of translation. The dominant model is that some codons or combinations of codons modulate the ribosome rate and thereby participate in the co-translational folding of proteins by separating the synthesis of defined parts of the protein (e.g. (50 , 51 , 52)). The availability of the tRNAs that decode synonymous codons, their requirement for wobble decoding interactions as well as the interactions between adjacent codons play fundamental roles in this model. In *Plasmodium*, both codon usage and the number of tRNA genes are very different from those of *E. coli* and thus make the expression of *Plasmodium* multi-domain proteins problematic in an *E. coli* heterologous expression system. Expression of the recombinant *Pf*-apiTyrRS directly from the *Plasmodium* mRNA sequence was indeed inefficient. Optimization of the synonymous codons implies the selection of optimal codons decoded only by abundant tRNAs and thus the simultaneous minimization of rare codons in the expression host (49, 53). This approach reduced, even more, the solubility of the produced *Pf*-apiTyrRS. Only, the use of harmonized codons, designed by the ANACONDA algorithm (22 , 23), allowed a better expression of the soluble protein. The major difference between both the Endo and the Opt genes and the Harm

gene is the systematic replacement of all abundant leucine codons (TTA, TTG, CTG, CTT and CTC) by the rarest leucine codon used in *E. coli* CTA ($3.85^0/_{00}$). Since the ribosome slows down upon encountering a rare codon it may help the protein domains to fold efficiently during translation, thereby the yield of soluble protein expression is increased. It confirms that recombinant genes containing a combination of fast and slow codons, matched to the codon usage of the heterologous host, facilitate co-translational protein folding and in our hands produced a biologically active *Pf*-apiTyrRS. This result suggests that such an approach could be used to overcome the difficulties encountered when expressing *Plasmodium* multi-domains proteins.

***Evolution of tyrosine identity***

Many experiments revealed that each tRNA holds specific sites involved in their correct recognition. Positions for positive or negative recognition participate in the correct aminoacylation or prevent false recognition, respectively (16). Experimental work on tyrosylation systems from different species allowed to establish the evolution of the tyrosine identity set (Figure 2C). The discriminator base A73 combined to anticodon bases G34 and U35 were recognized as common identity elements in all $tRNA^{Tyr}$ from eubacteria, archaebacteria, eukaryotes, and mitochondria (summarized in (20)). Base pair 1-72, at the extremity of the acceptor stem, is involved in all cytosolic tyrosylation systems whereas the long variable loop has been reported to be necessary for the correct tRNA recognition only for the bacterial TyrRSs (Figure 2A).

Based on our mutational analysis, the *Plasmodium* apicoplast tyrosylation system retains only one moderate (G34) and two weak (U35 and a long variable region) identity elements to ensure specific aminoacylation. Unlike other tyrosylation systems, this identity set does not involve any residue from the tRNA acceptor arm. Despite the conservation of the first A1-U72 base pair in all *Plasmodium* api-$tRNA^{Tyr}$, this position was not involved in tyrosylation, a situation that was already observed for the human mitochondrial tyrosine system (19). Unlike the long variable region in *Pf*-$apitRNA^{Tyr}$, the long variable region of *Pf*-$apitRNA^{Ser}_{GCU}$ potentially exhibits 8 base pairs and a specific tertiary interaction between the base pair $G_{45}$-$C_{48a}$ and $U_{20B}$ in the D-loop, determining its spatial orientation (33). Surprisingly, neither the size and structure nor the direction of the variable region are sufficient to hinder apicoplast tyrosylation, whereas, in *E. coli*, the simple insertion of the $tRNA^{Ser}$ variable region in $tRNA^{Tyr}$ is sufficient to decrease tyrosylation efficiency more than

300 fold (38). Moreover, mutations at identity positions 34, 35 and the removal of the long variable region lead to especially weak effects, with the highest loss in efficiency when position 34 is mutated reaching only 36, an unprecedented situation amongst tyrosylation systems.

Of the 26 other tRNA gene sequences encoded in the *Plasmodium* apicoplast genome, beyond *Pf*-apitRNA$^{Tyr}$, three tRNAs contain a G34U35 sequence (tRNA$^{Asn}_{GTT}$, tRNA$^{Asp}_{GTC}$, tRNA$^{His}_{GTG}$) (Figure S4) and one contains a G34 and a long variable region (tRNA$^{Ser}_{GCT}$). Thus, for avoiding the interactions of these tRNAs with the non-cognate *Pf*-apiTyrRS (e.g. *Pf*-apitRNASer is not a substrate for *Pf*-apiTyrRS, Table 1), they must display complementary structural features hindering tyrosylation. This type of negative discrimination has already been experimentally illustrated by locating antideterminants in tRNA molecules that prevent false recognition as it has been reviewed in (16). For example, in both *T. thermophilus* and *E. coli* tRNA$^{Tyr}$/TyrRS complexes, the long variable region in tRNA$^{Tyr}$ was shown to behave as a positive element but also to act as a negative determinant preventing recognition by heterologous synthetases (33 , 38).

As the principal conclusion of our analysis, we propose that the high A/T content of the *Plasmodium* apicoplast genome lowers the potential of identity combinations in apicoplast tRNAs significantly. In the specific case of *Pf*-apitRNA$^{Tyr}$, it led to the conservation of a minimal identity set with only 3 weak identity features that are positively recognized by the *Pf*-apiTyrRS. One may reasonably wonder if these elements are sufficient to drive tyrosylation in vivo efficiently and if tyrosylation specificity is essentially maintained because of the presence of negative determinants in other *Pf*-apitRNA that prevent mischarging by *Pf*-apiTyrRS. One can ask if this kind of recognition pattern is a general trend of apicoplast tRNAs or not.

**AKNOWLEDGMENTS**

We are grateful to Richard Giegé for critical reading of the manuscript. This work was supported by the CNRS and the Université de Strasbourg and the Fondation pour la Recherche Médicale (FRM) [grant numbers FDT201704337050]. This work has been published under the framework of the LABEX: ANR-10-LABX-0036_NETRNA.

**Table 1. Kinetic parameters of *Pf*-apiTyrRS for the aminoacylation of *Pf*-apitRNA$^{Tyr}$ and *Pf*-apitRNA$^{Ser}$ transcripts**

| | $K_m$ (μM) | S.D. $K_m$ | $k_{cat}$ ($10^{-3}sec^{-1}$) | S.D. $k_{cat}$ | $k_{cat}/K_m$ ($10^{-3}sec^{-1}.\mu M^{-1}$) | p-value | $^{a}$L (x-fold) |
|---|---|---|---|---|---|---|---|
| *E. coli* tRNA$^{Tyr}$ | 0,9 | 0,20 | 12 | 5.08 | 14,1 | 0.197 | 1.5 |
| *Pf*-apitRNA$^{Tyr}$ | 0.8 | 0.28 | 16 | 4 | 20.6 | | 1.0 |
| *Pf*-apitRNA$^{Ser}$ | nd | nd | | | - | - | >1000 |
| **Acceptor stem** | | | | | | | |
| $G_{73}$ | 1 | 0.29 | 9.8 | 5.37 | 9.2 | 0.021 | 2.2* |
| $C_{73}$ | 0.4 | 0.17 | 4 | 0.56 | 10.1 | 0.147 | 2.0 |
| $U_{73}$ | 0.6 | 0.10 | 3.1 | 0.07 | 4.9 | 0.019 | 4.2* |
| $G_1$-$C_{72}$ | 0.7 | 0.49 | 11.9 | 1.55 | 24.5 | 0.193 | 0.8 |
| $C_1$-$G_{72}$ | 0.5 | 0.25 | 5.8 | 1.06 | 11.7 | 0.007 | 1.8* |
| $C_2$-$G_{71}$/$U_3$-$A_{70}$ | 0.4 | 0.17 | 3 | 1.66 | 8.0 | 0.019 | 2.6* |
| **Anticodon triplet** | | | | | | | |
| $C_{34}$ | 6.2 | 3.45 | 3.2 | 1.58 | 0.6 | 0.014 | 36.0* |
| $G_{35}$ | 0.7 | 0.22 | 13.3 | 6.54 | 19.8 | 0.786 | 1.0 |
| $C_{35}$ | 1.9 | 0.03 | 4.9 | 1.62 | 2.6 | 0.012 | 7.9* |
| $A_{35}$ | 1.9 | 0.81 | 4.4 | 1.06 | 2.6 | 0.008 | 7.9* |
| $C_{36}$ | 0.7 | 0.35 | 3.5 | 1.13 | 5.8 | 0.015 | 3.6* |
| $U_{36}$ | 1.1 | 0.42 | 7.6 | 0.35 | 7.3 | 0.021 | 2.8* |
| $G_{36}$ | 1.6 | 0.14 | 11 | 0 | 6.9 | 0.026 | 3.0* |
| Ser ($C_{35}U_{36}$) | 1.3 | 0,35 | 1.3 | 0,36 | 1.1 | 0.014 | 19.1* |
| **Variable region (Vr)** | | | | | | | |
| ΔVr | 1.8 | 1.00 | 4.5 | 0.89 | 3.4 | 0.014 | 6.0* |
| SerVr | 0.8 | 0,42 | 5.6 | 2,30 | 7.6 | 0.013 | 2.7* |
| **D-loop (D-l)** | | | | | | | |
| $U_{20B}$ | 0.9 | 0.44 | 8.2 | 3.01 | 12.6 | 0.353 | 1.6 |
| SerVr+D-l | 0.9 | 0,21 | 8.5 | 2,12 | 10 | 0.047 | 2.1* |

$K_m$ and $k_{cat}$ values correspond to means arising from at least three independent experiments. The mean values of $k_{cat}/K_m$ and p-values were calculated independently. * significant at $p<0.05$. $^{a}$L values correspond to losses of the catalytic efficiency relative to wild-type *Pf*-apitRNA$^{Tyr}$. nd indicates not detectable. Results represent averages. Each data is an average of at least two independent experiments with the indicated error (S.D.)

## BIBLIOGRAPHY


1. Jackson KE, Habib S, Frugier M, Hoen R, Khan S, Pham JS, et al. Protein translation in Plasmodium parasites. Trends Parasitol. 2011;27(10):467-76.
2. Jackson KE, Pham JS, Kwek M, De Silva NS, Allen SM, Goodman CD, et al. Dual targeting of aminoacyl-tRNA synthetases to the apicoplast and cytosol in Plasmodium falciparum. Int J Parasitol. 2012;42(2):177-86.
3. Ralph SA, van Dooren GG, Waller RF, Crawford MJ, Fraunholz MJ, Foth BJ, et al. Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast. Nat Rev Microbiol. 2004;2(3):203-16.
4. Gardner M, Hall N, Fung E, White O, Berriman M, Hyman R, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 2002;419:498–511.
5. Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, et al. Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum. Science. 2003;299(5607):705-8.
6. Ibba M, Soll D. Aminoacyl-tRNA synthesis. Annu Rev Biochem. 2000;69:617-50.
7. Frugier M, Bour T, Ayach M, Santos MA, Rudinger-Thirion J, Theobald-Dietrich A, et al. Low Complexity Regions behave as tRNA sponges to help co-translational folding of plasmodial proteins. FEBS Lett. 2010;584(2):448-54.
8. Sharma A, Sharma A. Plasmodium falciparum mitochondria import tRNAs along with an active phenylalanyl-tRNA synthetase. Biochem J. 2015;465(3):459-69.
9. Eriani G, Delarue M, Poch O, Gangloff J, Moras D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature. 1990;347(6289):203-6.
10. Cusack S, Berthet-Colominas C, Hartlein M, Nassar N, Leberman R. A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 A. Nature. 1990;347(6290):249-55.
11. Huot JL, Enkler L, Megel C, Karim L, Laporte D, Becker HD, et al. Idiosyncrasies in decoding mitochondrial genomes. Biochimie. 2014;100:95-106.
12. Khan S. Recent advances in the biology and drug targeting of malaria parasite aminoacyl-tRNA synthetases. Malar J. 2016;15:203.
13. Pasaje CF, Cheung V, Kennedy K, Lim EE, Baell JB, Griffin MD, et al. Selective inhibition of apicoplast tryptophanyl-tRNA synthetase causes delayed death in Plasmodium falciparum. Sci Rep. 2016;6:27531.
14. Saint-Leger A, Sinadinos C, Ribas de Pouplana L. The growing pipeline of natural aminoacyl-tRNA synthetase inhibitors for malaria treatment. Bioengineered. 2016;7(2):60-4.
15. Giegé R, Springer M. Aminoacyl-tRNA Synthetases in the Bacterial World. EcoSal Plus. 2016;7(1).
16. Giegé R, Sissler M, Florentz C. Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. 1998;26(22):5017-35.
17. Giegé R, Eriani G. Transfer RNA recognition and aminoacylation by synthetases. eLS. 2014.

18. Wakasugi K, Quinn CL, Tao N, Schimmel P. Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant. EMBO J. 1998;17(1):297-305.
19. Bonnefond L, Frugier M, Giegé R, Rudinger-Thirion J. Human mitochondrial TyrRS disobeys the tyrosine identity rules. RNA. 2005;11(5):558-62.
20. Bonnefond L, Giegé R, Rudinger-Thirion J. Evolution of the tRNA(Tyr)/TyrRS aminoacylation systems. Biochimie. 2005;87(9-10):873-83.
21. Aurrecoechea C, Barreto A, Basenko EY, Brestelli J, Brunk BP, Cade S, et al. EuPathDB: the eukaryotic pathogen genomics database resource. Nucleic Acids Res. 2017;45(D1):D581-D91.
22. Pinheiro M, Afreixo V, Moura G, Freitas A, Santos MA, Oliveira JL. Statistical, computational and visualization methodologies to unveil gene primary structure features. Methods Inf Med. 2006;45(2):163-8.
23. Moura GR, Pinheiro M, Freitas A, Oliveira JL, Frommlet JC, Carreto L, et al. Species-specific codon context rules unveil non-neutrality effects of synonymous mutations. PLoS One. 2011;6(10):e26817.
24. Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, et al. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res. 2006;34(Web Server issue):W604-8.
25. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 2010;38(Web Server issue):W695-9.
26. Rost B, Yachdav G, Liu J. The PredictProtein server. Nucleic Acids Res. 2004;32(Web Server issue):W321-6.
27. Fechter P, Rudinger J, Giegé R, Theobald-Dietrich A. Ribozyme processed tRNA transcripts with unfriendly internal promoter for T7 RNA polymerase: production and activity. FEBS Lett. 1998;436(1):99-103.
28. Bonnefond L, Florentz C, Giegé R, Rudinger-Thirion J. Decreased aminoacylation in pathology-related mutants of mitochondrial tRNATyr is associated with structural perturbations in tRNA architecture. RNA. 2008;14(4):641-8.
29. Sohm B, Frugier M, Brule H, Olszak K, Przykorska A, Florentz C. Towards understanding human mitochondrial leucine aminoacylation identity. J Mol Biol. 2003;328(5):995-1010.
30. Chaudhry SR, Lwin N, Phelan D, Escalante AA, Battistuzzi FU. Comparative analysis of low complexity regions in Plasmodia. Sci Rep. 2018;8(1):335.
31. Kleeman TA, Wei D, Simpson KL, First EA. Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine. The Journal of biological chemistry. 1997;272(22):14420-5.
32. Bonnefond L, Frugier M, Touze E, Lorber B, Florentz C, Giegé R, et al. Crystal structure of human mitochondrial tyrosyl-tRNA synthetase reveals common and idiosyncratic features. Structure. 2007;15(11):1505-16.
33. Yaremchuk A, Kriklivyi I, Tukalo M, Cusack S. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. EMBO J. 2002;21(14):3829-40.

34. Sprinzl M, Vassilenko KS. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 2005;33(Database issue):D139-40.
35. Fechter P, Rudinger-Thirion J, Tukalo M, Giegé R. Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNA(Tyr) are conserved but expressed differently. Eur J Biochem. 2001;268(3):761-7.
36. Fechter P, Rudinger-Thirion J, Theobald-Dietrich A, Giegé R. Identity of tRNA for yeast tyrosyl-tRNA synthetase: tyrosylation is more sensitive to identity nucleotides than to structural features. Biochemistry. 2000;39(7):1725-33.
37. Himeno H, Asahara H, Tamura K, Hasegawa T, Ueda T, Watanabe K, et al. Discriminator base and anticodon as a tRNA identity element. Nucleic Acids Symp Ser. 1990(22):117-8.
38. Himeno H, Hasegawa T, Ueda T, Watanabe K, Shimizu M. Conversion of aminoacylation specificity from tRNA(Tyr) to tRNA(Ser) in vitro. Nucleic Acids Res. 1990;18(23):6815-9.
39. Hou YM, Schimmel P. Modeling with in vitro kinetic parameters for the elaboration of transfer RNA identity in vivo. Biochemistry. 1989;28(12):4942-7.
40. Bedouelle H. Recognition of tRNA(Tyr) by tyrosyl-tRNA synthetase. Biochimie. 1990;72(8):589-98.
41. Sherman JM, Rogers K, Rogers MJ, Soll D. Synthetase competition and tRNA context determine the in vivo identify of tRNA discriminator mutants. J Mol Biol. 1992;228(4):1055-62.
42. Messmer M, Gaudry A, Sissler M, Florentz C. Pathology-related mutation A7526G (A9G) helps in the understanding of the 3D structural core of human mitochondrial tRNA(Asp). RNA. 2009;15(8):1462-8.
43. Celis JE, Hooper ML, Smith JD. Amino acid acceptor stem of E. coli suppressor tRNA tyr is a site of synthetase recognition. Nat New Biol. 1973;244(139):261-4.
44. Lee CP, Mandal N, Dyson MR, RajBhandary UL. The discriminator base influences tRNA structure at the end of the acceptor stem and possibly its interaction with proteins. Proc Natl Acad Sci U S A. 1993;90(15):7149-52.
45. Bour T, Akaddar A, Lorber B, Blais S, Balg C, Candolfi E, et al. Plasmodial aspartyl-tRNA synthetases and peculiarities in Plasmodium falciparum. The Journal of biological chemistry. 2009;284(28):18893-903.
46. Hoen R, Novoa EM, Lopez A, Camacho N, Cubells L, Vieira P, et al. Selective inhibition of an apicoplastic aminoacyl-tRNA synthetase from Plasmodium falciparum. Chembiochem. 2013;14(4):499-509.
47. Mailu BM, Ramasamay G, Mudeppa DG, Li L, Lindner SE, Peterson MJ, et al. A nondiscriminating glutamyl-tRNA synthetase in the plasmodium apicoplast: the first enzyme in an indirect aminoacylation pathway. The Journal of biological chemistry. 2013;288(45):32539-52.
48. Pham JS, Sakaguchi R, Yeoh LM, De Silva NS, McFadden GI, Hou YM, et al. A dual-targeted aminoacyl-tRNA synthetase in Plasmodium falciparum charges cytosolic and apicoplast tRNACys. Biochem J. 2014;458(3):513-23.
49. Hu S, Wang M, Cai G, He M. Genetic code-guided protein synthesis and folding in Escherichia coli. The Journal of biological chemistry. 2013;288(43):30855-61.

50. Thanaraj TA, Argos P. Protein secondary structural types are differentially coded on messenger RNA. Protein Sci. 1996;5(10):1973-83.
51. Thanaraj TA, Argos P. Ribosome-mediated translational pause and protein domain organization. Protein Sci. 1996;5(8):1594-612.
52. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat Struct Mol Biol. 2009;16(3):274-80.
53. Grosjean H, Fiers W. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene. 1982;18(3):199-209.
54. Chaubey S, Kumar A, Singh D, Habib S. The apicoplast of Plasmodium falciparum is translationally active. Mol Microbiol. 2005;56(1):81-9.

## FIGURE LEGENDS

**Figure 1. Expression of a soluble and active *Pf*-apiTyrRS in *E. coli***

(**A**) The overall organization of *Pf*-apiTyrRS compared with TyrRSs from human mitochondria, eubacteria, archaea and, eukarya. Each structural domain is given in a specific color: blue for mitochondrial and apicoplast targeting signals, black for the catalytic domains, green for anticodon-binding domains and grey for C-terminal additional domains (S4-like in eubacteria and organelles and EMAPII-like in vertebrates (31). Red dots indicate sequences specific for class I aaRSs present in the catalytic domain. *Plasmodium*-specific insertions found in *Pf*-apiTyrRS are shown in yellow. (**B**) Western blot and the corresponding coomassie-stained gel showing *Pf*-apiTyrRS expressed either from the *Plasmodium* gene (Endo), the optimized gene sequence for *E. coli* expression (Opt) and the harmonized gene sequence (Harm). T stands for Total extract, and S stands for Soluble extract (supernatant of the centrifuged total extract). On the coomassie-stained gel, differences in pattern between lines are indicated with asterisks. The green asterisk designates the band corresponding to the overexpression of the Harm *Pf*-TyrRS and red asterisks indicate additional bands that could be degradations (**C**) Quantification and comparison of the purification yields of *Pf*-apiTyrRS expressed from Endo, Opt and Harm genes. Results are an average of two independent experiments. (**D**) Aminoacylation of *E. coli* $tRNA^{Tyr}$ (0.8 μM) performed with *Pf*-apiTyrRSs (100 nM) produced from the 3 gene versions. Results are an average of three independent experiments.

**Figure 2. Sequence of *Pf*-apit$RNA^{Tyr}$ and evolutionary view of the tyrosine identity**

(**A**) Comparison of tyrosine-specific apitRNAs belonging to different *Plasmodium* species. The features responsible for the cloverleaf-fold are indicated on the top: Acc-s stands for acceptor stem, D-s for D-stem, D-l for D-loop, Ac-l for anticodon-loop, T-s for T-stem and T-l for T-loop. Some structural elements or nucleotides involved in specific tertiary interactions are highlighted in red. (**B**) Secondary structure of *Pf*-apit$RNA^{Tyr}$. The architectural organization of the D-loop (α and β domains on both sides of residues G18G19) is indicated. Notice the non-canonical base pair G13-A22 located at the end of the D-arm. The tRNA is numbered according to (34) .(**C**) Schematic 2D structures of $tRNA^{Tyr}$ showing the residues involved in tyrosylation in different phylae (46, 19 , 20 , 35 , 36 , 37 , 38, 39 , 40 , 41 ). The residues involved in tyrosine identity are explicitly given in uppercase. Strengths of the tyrosine identity elements are indicated with colors: red (loss in catalytic activity >100),

orange (loss comprised between 10 and 100) and green (loss comprised between 5 and 10). Lowercase letters are given to highlight the conservation of the residues in the anticodon triplet and at positions 1-72 and 73 in the acceptor stem despite their exclusion from the identity set. The question mark shows that position 36 was not tested in eubacteria.

**Figure 3. Probing of *Pf*-apitRNA$^{Tyr}$ transcript**

(**A**) Autoradiograms corresponding to probing experiments on 5'-labeled *Pf*-apitRNA$^{Tyr}$ transcript with lead (left panel) and RNases V1, S1 and nuclease T1 (right panel). The final concentrations of PbOAc are indicated on the top of the autoradiogram. The G lane corresponds to a denaturing RNase T1 ladder and lane HA to an alkaline ladder. C lane indicates the control experiment performed without probes. The cuts are summarized in *Pf*-apitRNA$^{Tyr}$ cloverleaf structure. The tRNA is numbered according to (34). Location of D-, anticodon- and T-loops as well as of the large variable region is given. Unprobed nucleotides at the 3' and 5'-ends are indicated by dotted lines on the tRNA structure. (**B**) Autoradiogram corresponding to footprinting experiments: the 5'-labeled *Pf*-apitRNA$^{Tyr}$ transcript was incubated with S1 and T1 nucleases in the absence (-) or presence (+) of *Pf*-apiTyrRS. Controls (C) were done without nucleases. G and HA indicate T1 and alkaline ladders, respectively. Location of D-, anticodon- and T-loops as well as of the large variable region is given. *Pf*-apitRNA$^{Tyr}$ residues that are protected from nucleases in the presence of the enzyme are indicated in orange (strong protections) and yellow (weak protections).

**Figure 4: Schematic 2D structures of tRNA$^{Tyr}$ highlighting the location and the nature of the mutated residues**. Mutated sequences are explicitly given; The variable region is either deleted of 10 residues (ΔVr: from 47 to 47i) or replaced with the variable region of *Pf*-apitRNA$^{Ser}$ (SerVr). Similarly, the *Pf*-apitRNA$^{Tyr}$ D-loop was replaced with the *Pf*-apitRNA$^{Ser}$ D-loop.

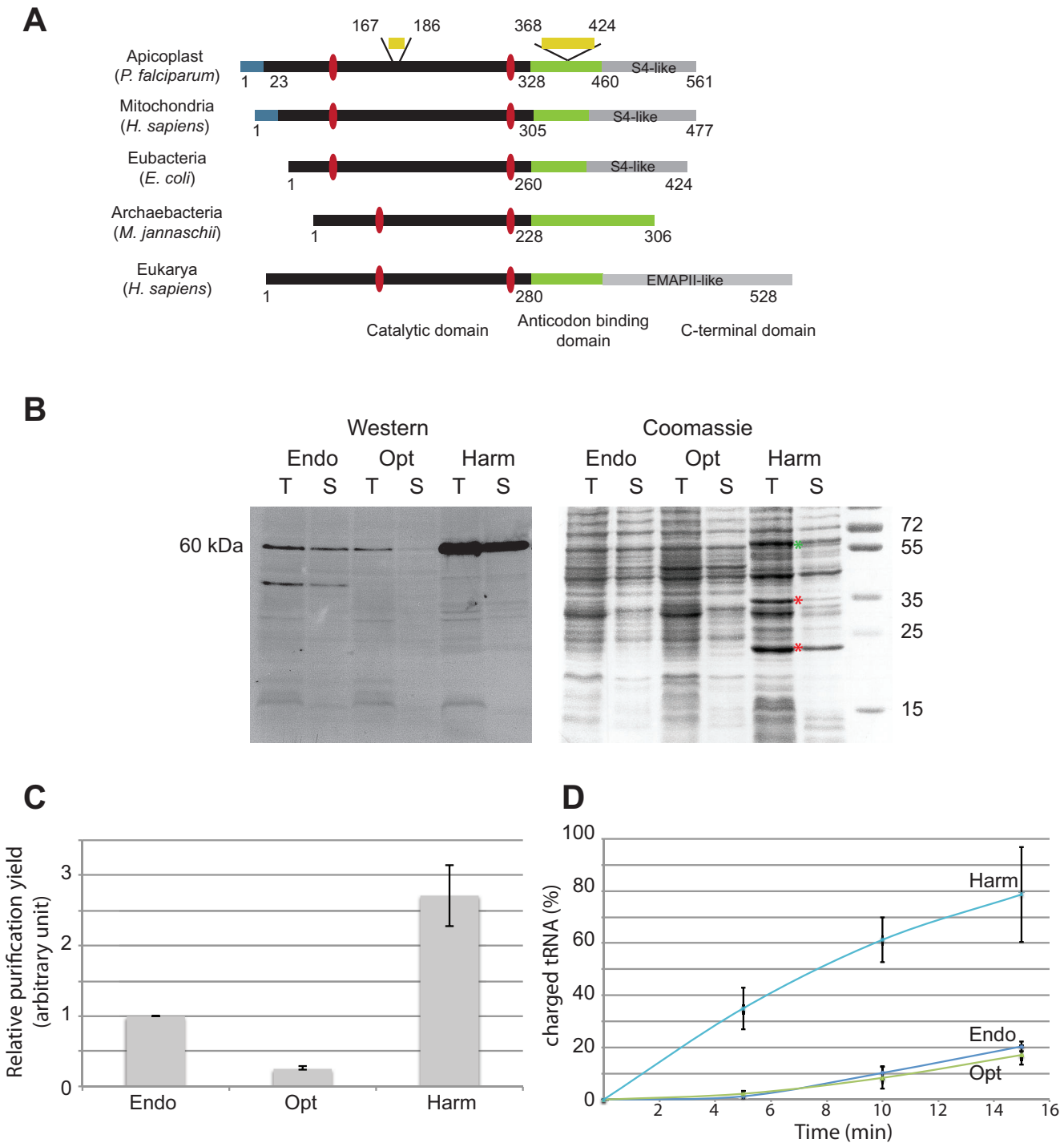

A
167
186
368
424
Apicoplast
(P. falciparum)
1
23
328
460
561
S4-like
Mitochondria
(H. sapiens)
1
305
477
S4-like
Eubacteria
(E. coli)
1
260
424
S4-like
Archaebacteria
(M. jannaschii)
1
228
306
Eukarya
(H. sapiens)
1
280
528
EMAPII-like
Catalytic domain
Anticodon binding domain
C-terminal domain
B
Western
Coomassie
Endo
Opt
Harm
T S T S T S
60 kDa
72
55
35
25
15
C
Relative purification yield (arbitrary unit)
0
1
2
3
Endo
Opt
Harm
D
charged tRNA (%)
0
20
40
60
80
100
Harm
Endo
Opt
2
4
6
8
10
12
14
16
Time (min)

**Figure 1**

**A**

```
                5'     Acc-s  D-s     D-l   D-s  Ac-s  Ac-l  Ac-s  variable region  T-s   T-l   T-s   Acc-s   3'
                        >>>>>>> >>>>  α3   β4   <<<< >>>>>       <<<<<                 >>>>>       <<<<< <<<<<<<
P.gallinaceum          -AAGTTAATGCCTG.AGTGGTTAA.AAGGAATGGACTGTAAATCCAT.TGATTATATATCTAC.ATCAGTTCAAATCTGAT.TTAACTT.A
P.malariae             -AAGTTAATGCCTG.AGAGGTTAA.AAGGAATGGACTGTAAATTCAT.TGATTAAATATCTAC.ATCAGTTCAAATCTGAT.TTAACTT.A
P.falciparum           -AAGTTAATGCCTG.AGTGGTTAA.AAGGAATGGACTGTAAATCCAT.TGATAATATATCTAC.ATCAGTTCAAATCTGAT.TTAACTT.A
P.reichnowi            -AAGTTAATGCCTG.AGTGGTTAA.AAGGAATGGACTGTAAATCCAT.TGATAATATATCTAC.ATCAGTTCAAATCTGAT.TTAACTT.A
P.knowlesi             -AAGTTAATGCCTG.AGTGGTTAA.AAGGAATGGACTGTAAATCCAT.TGATTATATATCTAC.ATCAGTTCAAATCTGAT.TTAACTT.A
P.fragile              -AAGTTAATGCCTG.AGTGGTTAA.AAGGAATGGACTGTAAATCCAT.TGATTATATATCTAC.ATCAGTTCAAATCTGAT.TTAACTT.A
P.gaboni               AAAGTTAATGCCTG.AGTGGTTAA.AAGGAATGGACTGTAAATCCAT.TGATTAAATATCTAC.ATCAGTTCGAATCTGAT.TTAACTT.-
P.chabaudi             -AAGTTAATGCCTG.AGCGGTTAA.AAGGAATGGACTGTAAATCCAT.TGGTTTTATATCTAC.ATCAGTTCGAATCTGAT.TTAACTT.A
P.yoelli               -AAGTTAATGCCTG.AGTGGTTAA.AAGGAATGGACTGTAAATCCAT.TGGTGTTATACCTAC.ATCAGTTCAAATCTGAT.TTAACTT.A
P.berghei              -AAGTTAATGCCTG.AGGGGTTAA.AAGGGATGGACTGTAAATCCAT.TGGTTTTATATCTAC.ATCAGTTCAAATCTGAT.TTAACTT.A
                        ************* ** ****** **** ************* *** ** *   *** **** ******** ******** *******
                                    13      20B 22                                  47i
```

**B**

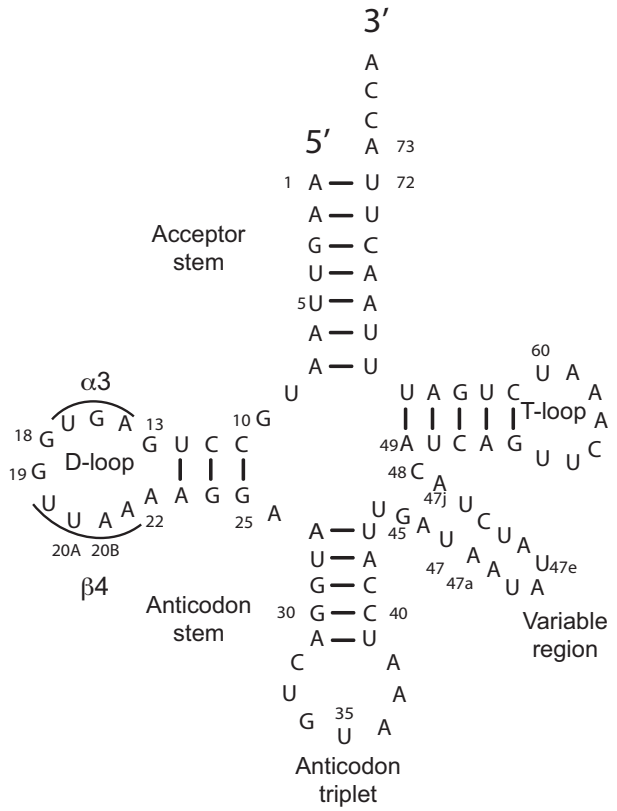



**Figure 2**

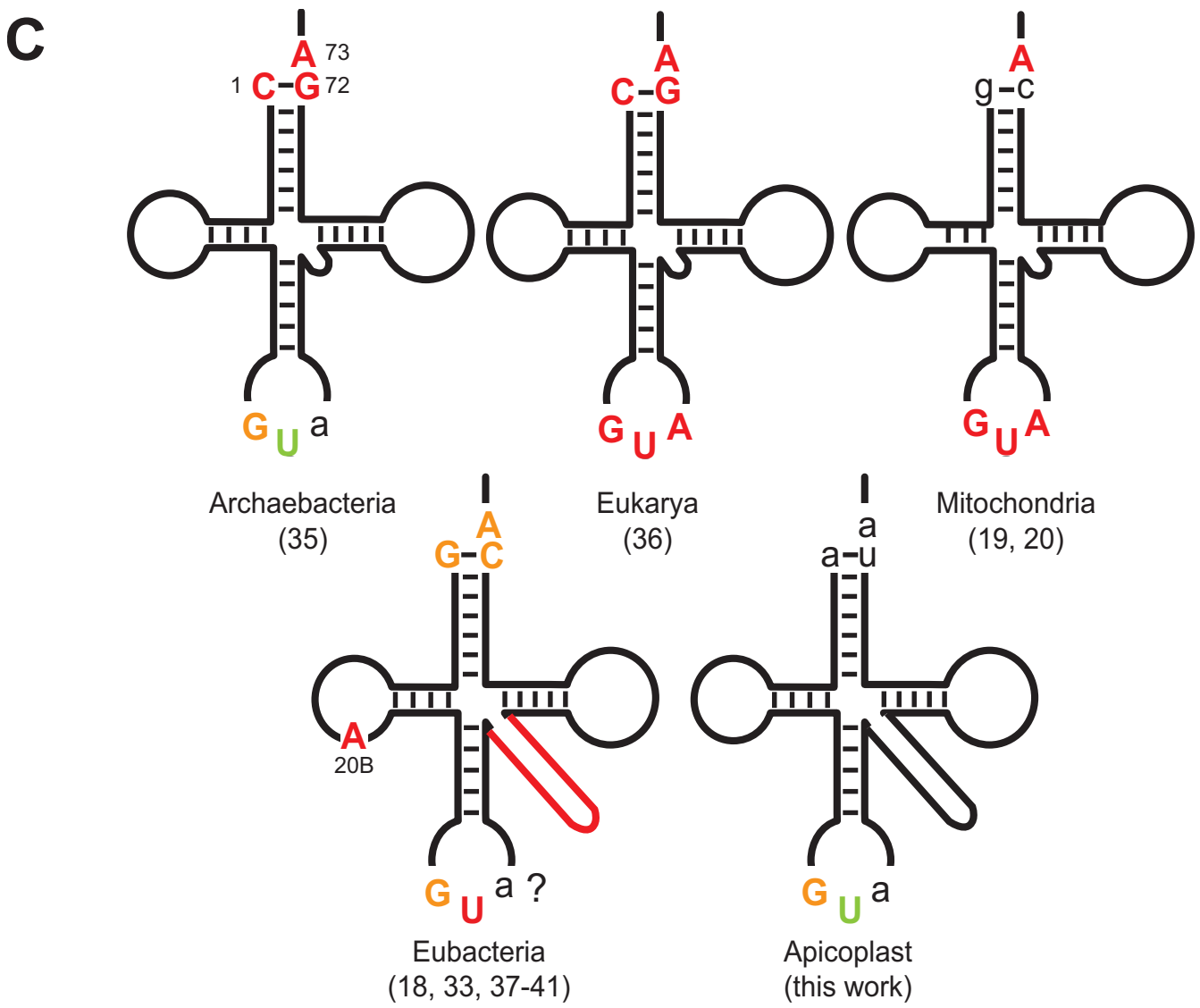

C
A 73
1 C–G 72
G U a
Archaebacteria
(35)
A
C–G
G U A
Eukarya
(36)
A
g–c
G U A
Mitochondria
(19, 20)
A
G–C
A
20B
G U a ?
Eubacteria
(18, 33, 37-41)
a
a–u
G U a
Apicoplast
(this work)

**Figure 2**

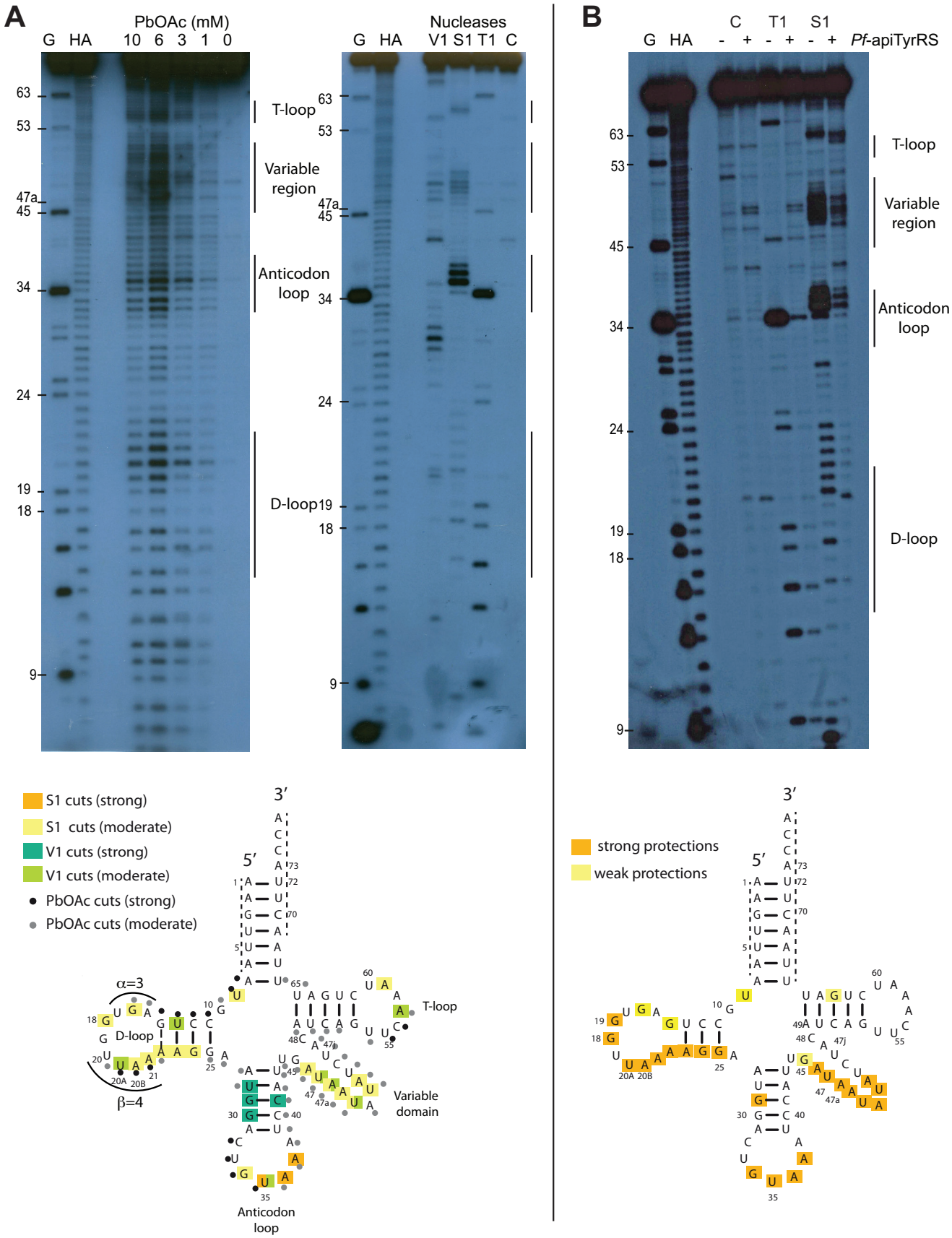

A
PbOAc (mM)
G HA 10 6 3 1 0
Nucleases
G HA V1 S1 T1 C
63
53
47a
45
34
24
19
18
9
T-loop
Variable region
Anticodon loop
D-loop
S1 cuts (strong)
S1 cuts (moderate)
V1 cuts (strong)
V1 cuts (moderate)
PbOAc cuts (strong)
PbOAc cuts (moderate)
3′
5′
α=3
β=4
D-loop
T-loop
Variable domain
Anticodon loop
B
C T1 S1
G HA - + - + - + Pf-apiTyrRS
T-loop
Variable region
Anticodon loop
D-loop
strong protections
weak protections

**Figure 3**

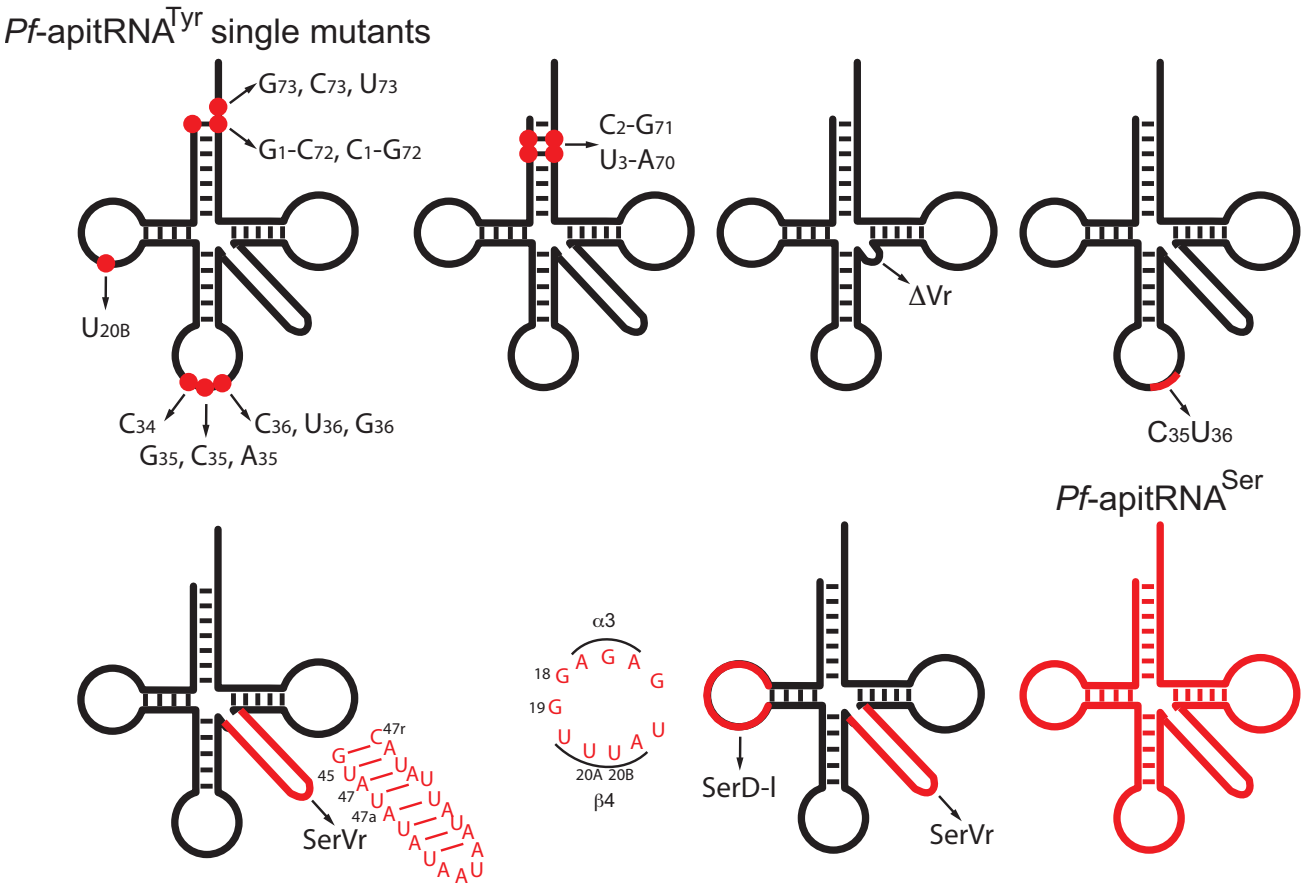

*Pf*-apitRNA$^{Tyr}$ single mutants
G73, C73, U73
G1-C72, C1-G72
U20B
C34
G35, C35, A35
C36, U36, G36
C2-G71
U3-A70
ΔVr
C35U36
*Pf*-apitRNA$^{Ser}$
SerVr
47r
45
47
47a
α3
18
19
20A 20B
β4
SerD-l
SerVr

**Figure 4**

## SUPPORTING INFORMATION CAPTIONS

**Figure S1. Multiple sequence alignments of *Plasmodium* apicoplast TyrRSs with the human mito-TyrRS**

Initiation of recombinant *Pf*-apiTyrRS is indicated in cyan. The color code follows that of Figure 1A: residues belonging to the catalytic domain are in black with class I signature motifs highlighted in red, residues from the anticodon-binding domain are in green, the S4-like domain in grey and the two *Plasmodium*-specific insertions are colored in yellow. Alignments were performed with Tcoffee (http://tcoffee.crg.cat/apps/tcoffee/do:expresso) (24) and predicted β-sheets and α-helices (http://www.predictprotein.org/) (21) of *Pf*-apiyRS are indicated with green arrows and red rectangles, respectively.

**Figure S2. DNA sequences encoding *Pf*-apiTyrRS**

Alignment of nucleotide sequences corresponding to "endogenous" (Endo), "optimized" (Opt) and "harmonized" (Harm) gene sequences coding for *Pf*-apiTyrRS. The amino acid sequence of the protein is given in bold on the top. All six Leucine (L) residues were highlighted: in red when the codon usage in *E. coli* higher than $10^0/_{00}$ (TTG, TTA, CTG, CTT and CTC) and in green for CTA ($3.85^0/_{00}$).

(54)(https://openwetware.org/mediawiki/index.php?title=Escherichia_coli/Codon_usage&oldid=820473).

**Figure S3. ATP/PPi exchange assays**. Experiments were performed according to (Campanacci et al. 2004, JMB), in the presence of tyrosine (2 mM) and radiolabeled [$^{32}$P]PPi (20 cpm/pmol, this high specific activity was used in order to be able to detect even low exchange activity) and 0.8 µM of both the full-length *Pf*-apiTyrRS (squares) and the *Pf*-apiTyrRS deprived of the S4-like domain (spheres). A negative control with no tyrosine in the sample was done in parallel (triangles).

**Figure S4. Sequences of *Plasmodium* apicoplast tRNAs**

Structural alignment of apitRNAs from *P. falciparum*. Accession numbers are given, and the anticodon sequence is indicated in red. Notice the presence of introns in both $tRNA^{Leu}_{TAA}$ and potentially in $tRNA^{Met}_{CAT}$-2 gene sequences. Based on their sequences, we could not determine which of the two $tRNA^{Met}_{CAT}$ is initiator or elongator.

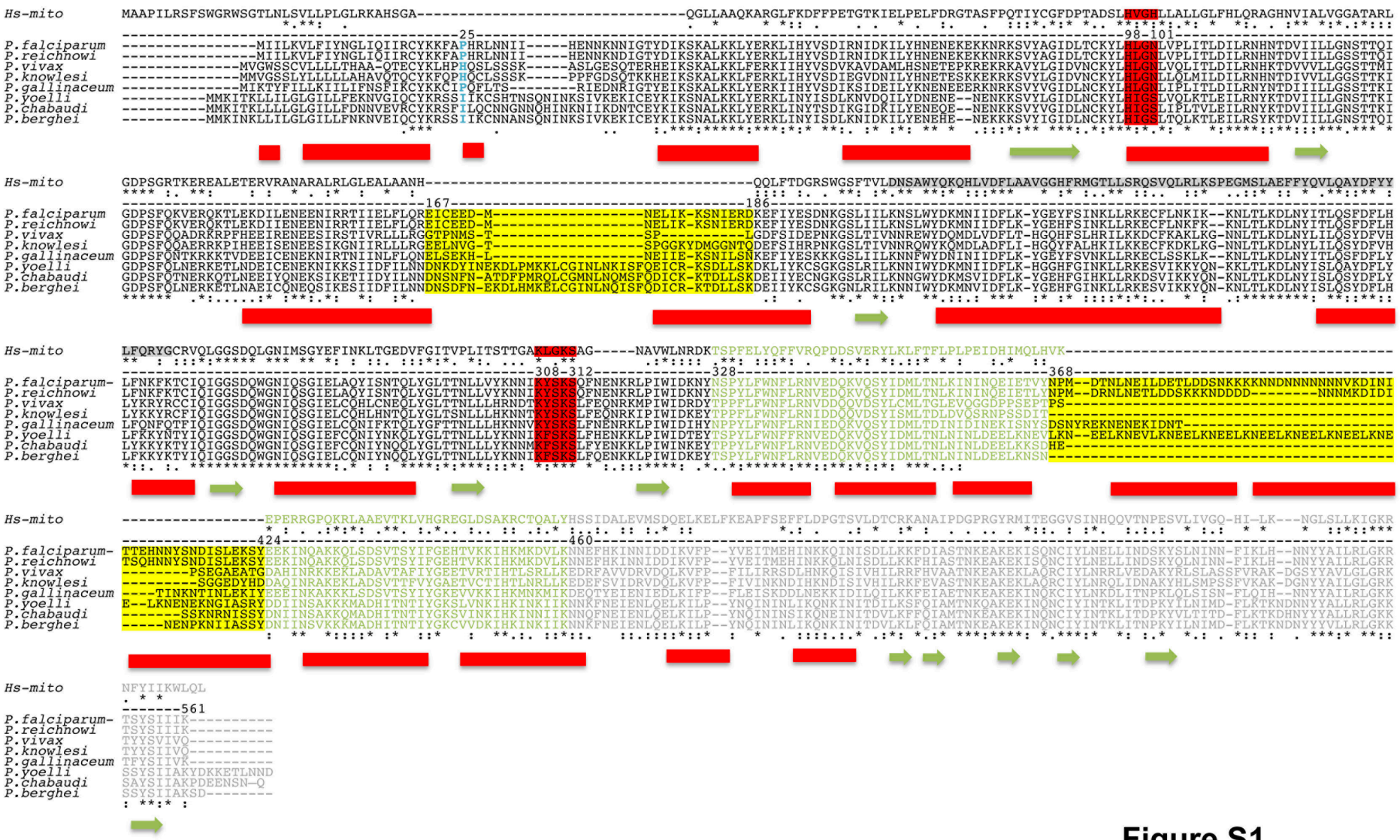

Hs-mito
P.falciparum
P.reichnowi
P.vivax
P.knowlesi
P.gallinaceum
P.yoelli
P.chabaudi
P.berghei
P.falciparum-
25
98-101
167
186
308-312
328
368
424
460
561

**Figure S1**

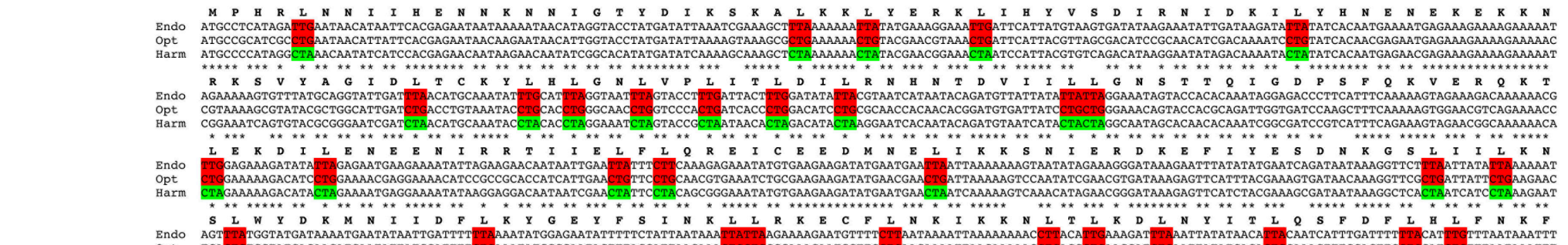



**Figure S2**

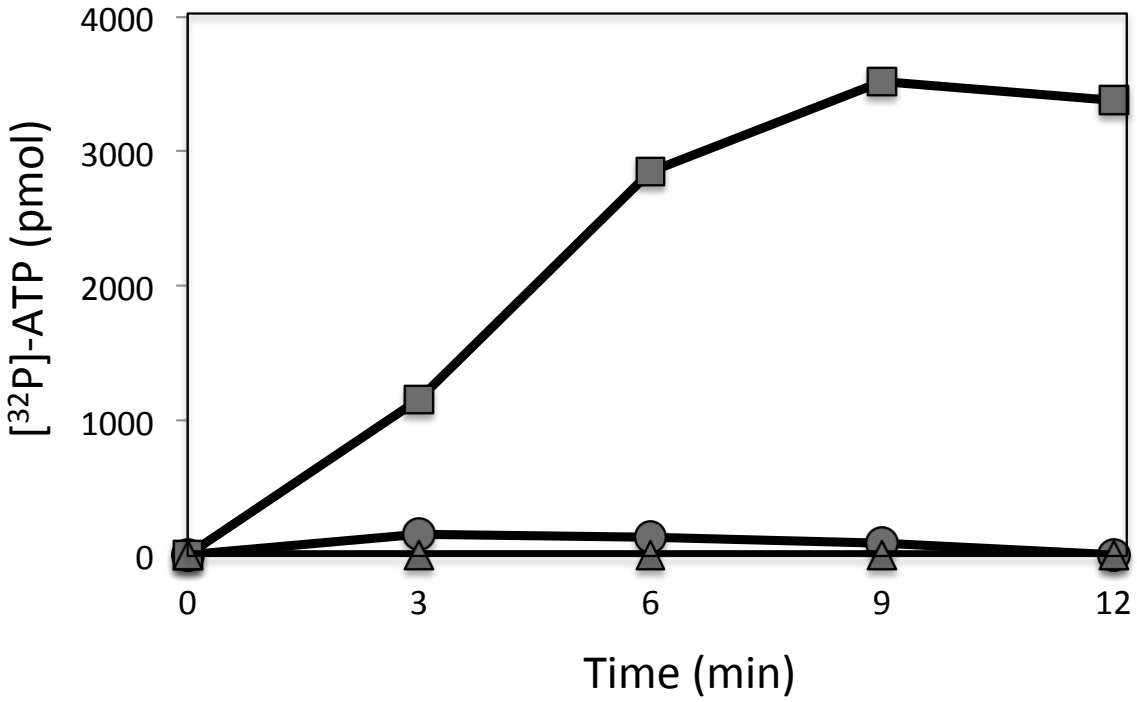



**Figure S3**

```
                                 5'      Acc-s     D-s      D-l       D-s  Ac-s  Ac-l  Ac-s      variable region      T-s   T-l   T-s   Acc-s   3'
                                         >>>>>>>  >>>>               <<<< >>>>>        <<<<<                           >>>>>        <<<<< <<<<<<<
AlaTGC (PF3D7_API05600)                  GGGAATATAGTTT....AATGGTA...AAATCTTATTTTTGCATAATAA.AGAT....................AGTAGTTCAATTCTACT.TATTTCC.A
ArgACG (PF3D7_API05100)                  AAACTTGTAATTT....AATGGATA..AAATATATAAATACGAATTATA.AAAT....................AAAAGTTCAATTCTTTT.CAAGTTT.A
ArgTCT (PF3D7_API05300)                  AAATCTATAATTT....AATGGATA..AAATAAAAACCTTCTAAGTTTT.ATAT....................GTAAGTTCAAATCTTAC.TAGATTT.A
AsnGTT (PF3D7_API05500)                  TTCTTAATAGCTT....AGTGGTTA..AAGCATTCGGCTGTTAACCGAA.ATAC....................ACTAGTTCAATTCTAGT.TTAAGAA.G
AspGTC (PF3D7_API00900)                  GAATTTGTTAGTT.TAATTAGGTAA..AAATATTATTTTGTCATAATAA.AGAAT...................ACGAGTTCAATTCTCGT.CAAATTC.G
CysGCA (PF3D7_API00300)                  AATGATATAACTT....AATTGATA..AAGTAAATAATTGCAAATTATT.ATA.....................TTTCAGTTTGAATCTGA.ATATCAT.T
GlnTTG (PF3D7_API03200)                  TAGAATATAACCA....AAAGGTTA..AGGTAATGAATTTTGATTTCAT.TAAT....................ATAGGTTCGAATCCTAT.TATTCTA.A
GluTTC (PF3D7_API01100)                  ACTTTTATCGTTT....AAAGGTA...AGACATCTTTTTTTCAAGAAGA.AAA.....................TAGGAATTCGATTTTCC.TTAAAAG.T
GlyACC (PF3D7_API03300)                  ATGAATATAATTT....AATGATA...AAATACAATTTTACCATAATTG.TTA.....................TAAGAGTTTGAATCTCT.TTATTCA.T
GlyTCC (PF3D7_API03700)                  ACAAATATAGTTT...AATCGGTA...AAATATTAATTTTCCAAATTAA.TGAT....................ATGGATTCAATTTCCAT.TATTTGT.A
HisGTG (PF3D7_API00200)                  ATAAATATAATCT....AATGGTTA..AGATGAAGAATTGTGGTTTCTT.TTAT....................ATGAGTTCAAATCTCTT.TATTTAT.C
IleGAT (PF3D7_API05800)                  AGGTTTTTAGTTT....AATGGTTA..AAACATACTCTTGATAAGGGTA.AAAT....................TTTAGTTCAATTCTAAA.AAAACCT.A
LeuTAG (PF3D7_API05400)                  ATGAATATGGCGA...AATAGGTAA..ACGCACTAAATTTAGATTTTAG.TTATTATAAT..............AAGAGTTCAAATCTCTT.TATTCAT.A
LeuTAA Pf3D7_API_v3:809..1,028-intron    AGAGATATGGTGA..AATTTGGTAT..ACACAATGGACTTAAAATCCAT.TAACATTATTGTTGT.........AAGGGTTCAAATCCCTT.TATCTCT.A
LysTTT (PF3D7_API01000)                  GAATTACTAGCTT...AATTGGTA...GAGTACTCGACTTTTAATCGAA.TGGTT...................CTGAGTTCAAATCTCAG.GTAGTTC.A
MetCAT-1(PF3D7_API05000)                 AGCGAAATAGAGC...ATAAGGAAA..GTTCGTCGGATTCATGCTCCGA.AGGTA...................ATCGGTTCAATTCCGTT.TTTCGCT.T
MetCAT-2(PF3D7_API00600)-intron ?        ACATTTATAGCTA....AGTGGTCGA.AAGCAATGGACTCATAATTGAT.CATC....................AGTAGTTCGAATCTACT.TAAATGT.A
PheGAA (PF3D7_API03100)                  GTCATAATAGCTC....AATGGTA...GAGCAATGGATTGAAGATCCAT.GTGTT...................ATCAGTTCAAATCTGAT.TTTTGAC.A
ProTGG (PF3D7_API01200)                  CAGAATATAGTGT....AATGGTA...ACATATCTATTTTGGGAATAGA.AGAAT...................ATAGGTTCAAATCCTAT.TTTTCTG.A
SerTGA (PF3D7_API03900)                  AGAGAAATGACAG....AGTGGTTTA.TTGTGTTTGATTTGAGATCAAA.AAAATATAAATATATTTC......ATGGGTTCAAATCCCAT.TTTCTCT.T
SerGCT (PF3D7_API00800)                  AGAGAAATGACTG....AGAGGTTTA.TAGTTATAAATTGCTAATTTAT.TGTATATATAATAATATTATACC.AAGGGTTCGAATCCCTT.TTTCTCT.A
ThrTGT (PF3D7_API04800)                  GCTAAAGTAGCTT...AATTGGTAA..AGCA.ACTGATTTGTAATCAGT.AGATT...................ATGAGTTCAAATCTCAC.CATTAGC.T
ThrTGT (PF3D7_API04800)                  GCTAAAGTAGCTT...AATTGGTA...AAGCAACTGATTTGTAATCAGT.AGATT...................ATGAGTTCAAATCTCAC.CATTAGC.T
TrpCCA (PF3D7_API03400)                  ATGTCTTTAATTT....AAAGTAA...AAATATAAATTTCCAAAATTTA.TAAT....................AAAGGTTCGAATCCTTT.AGGACAT.G
TyrGTA (PF3D7_API00700)                  AAGTTAATGCCTG....AGTGGTTAA.AAGGAATGGACTGTAAATCCAT.TGATAATATATCTAC.........ATCAGTTCAAATCTGAT.TTAACTT.A
ValTAC (PF3D7_API05200)                  AAGTAATTAACTT....AGAGGTA...AAGTTTCTGCTTTACATACAGA.AGACC...................ATTGGTTCGAATCCAAT.ATTACTT.A
```

**Figure S4**

## II. LA SELECTIVITE DE TRIP *IN VITRO*

## 1. Le MIST

### *1.1. Principe du MIST*

La technique MIST, ou Microarray Identification of Shifted tRNAs, a été développée par le Dr. Renaud Geslain (Eriani, *et al.*, 2015). Cette technique permet d'identifier dans une population les ARNt qui ont le plus d'affinité pour une protéine donnée. Cette approche a été adaptée à partir du protocole développé dans l'équipe du Dr. Tao Pan en 2006 (Dittmar, *et al.*, 2006). Ce protocole permet de quantifier les ARNt cellulaires grâce à des « tRNA microarrays », ou puces à ARNt. Chaque puce est composée de différentes sondes d'ADN complémentaires à chaque séquence d'ARNt. Les ARNt présents dans un échantillon sont marqués radioactivement, et hybridés à la puce. Ainsi la proportion de chaque ARNt dans l'échantillon peut être déterminée grâce à l'intensité de la radioactivité retenue sur la puce. Dans le cas de l'étude de l'expression des ARNt dans différents tissus humains (Dittmar, *et al*., 2006), les sondes ont été conçues pour s'hybrider à la séquence complète de l'ARNt (70 - 80 nucléotides). Comme montré dans l'article, la longueur des sondes augmente considérablement l'efficacité d'hybridation et permettent de distinguer deux ARNt qui ne diffèrent que par dix nucléotides.

Le Dr. Renaud Geslain a adapté cette technique à l'étude de complexes formés entre ARNt et protéine (Eriani, *et al*., 2015). Pour cela, il a ajouté une étape de sélection avant l'identification sur puce à ARNt. Dans cette étape initiale, les complexes ARNt/protéine formés sont purifiés sur un gel natif d'électrophorèse, les ARNt sont élués, marqués puis analysés.

### *1.2. Purification de la protéine Pf*tRip

La protéine *Pf*tRip est purifiée selon trois étapes de chromatographie liquide de haute performance (HPLC). La première étape (Figure 22A) est une chromatographie d'affinité sur colonne de nickel, NiNTA. Elle permet d'éliminer la plupart des protéines contaminantes, ainsi que les acides nucléiques liés à la protéine. La deuxième étape est une colonne d'échange d'anions (UnoQ), qui va séparer la forme active de *Pf*tRip des formes inactives qui

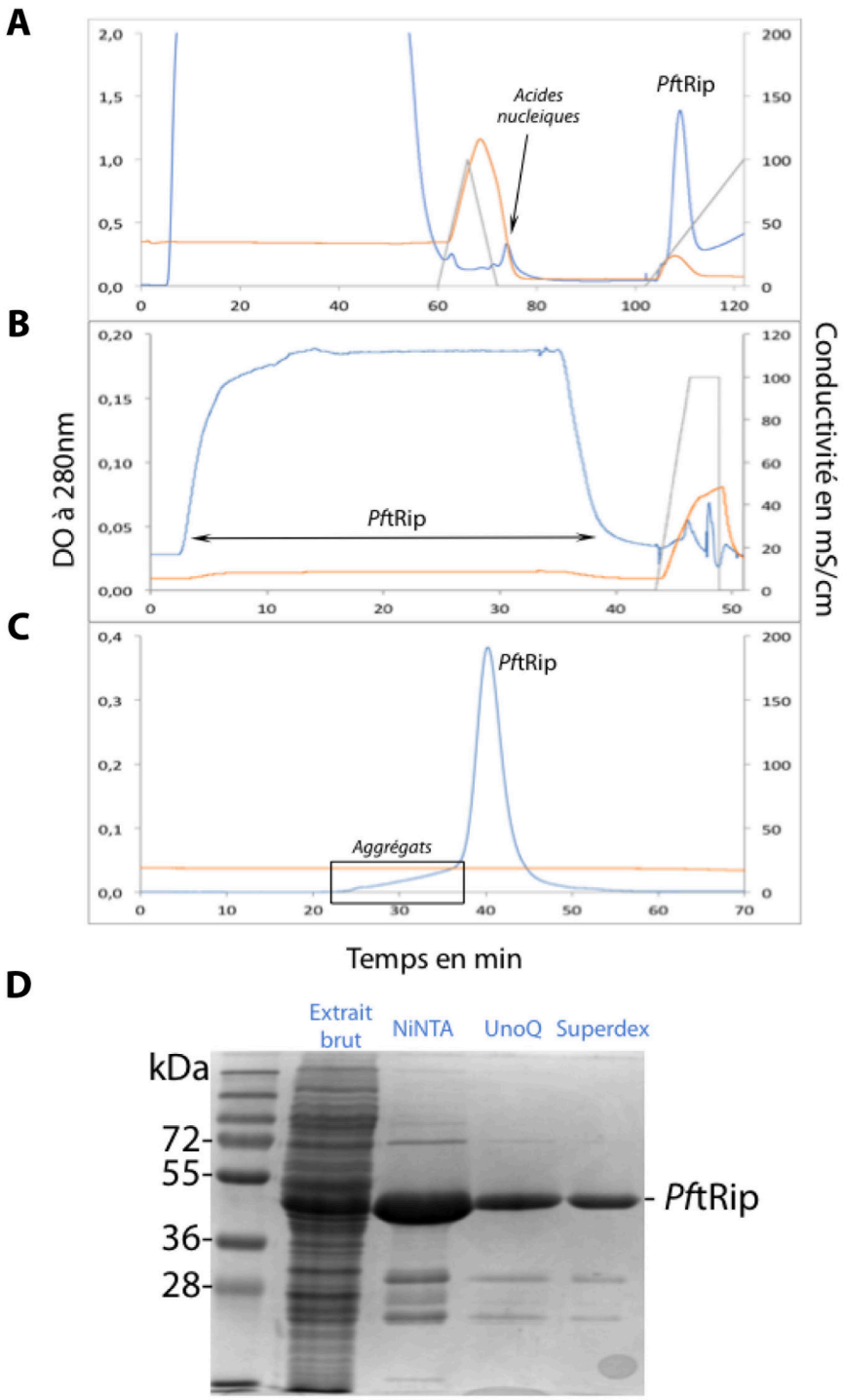



**Figure 22 : Les différentes étapes de purification de *Pf*tRip recombinante : (A)** Chromatographie sur colonne NiNTA. Le protocole du fournisseur a été adapté à la purification d'une protéine liant les ARN, c'est à dire qu'un gradient de sels a été ajouté. Celui-ci permet de décrocher les acides nucléiques liés sur *Pf*tRip. **(B)** Chromatographie sur colonne UnoQ. Seule la protéine active ne se fixe pas sur la résine. **(C)** Chromatographie sur colonne Superdex 200 permet d'éliminer les derniers contaminants. Le suivi de la conductivité est indiqué en rouge, la densité optique en bleu, et les gradients théoriques en noir. **(D)** Analyse sur gel PAGE-SDS 12% de *Pf*tRip après chaque étape de purification.

restent liées à la résine (Figure 22B). Enfin, les aggrégats de *Pf*tRip et les protéines qui contaminent encore l'échantillon sont éliminés par chromatographie d'exclusion de taille sur une colonne Superdex 200 (Figure 22C). Selon ce protocole, entre 3 et 4 mg de *Pf*tRip pure et active sont obtenus à partir de 250 mL de culture.

La Figure 22D montre le résultat après chaque étape du protocole de purification. Après la dernière colonne, trois bandes sont encore visibles, la plus intense, correspondant à un monomère de *Pf*tRip (50 kDa). Les deux autres bandes ont été identifiées par spectrométrie de masse et correspondent aux fragments C- et N-terminaux de *Pf*tRip. Cependant, leur concentration est négligeable par rapport à celle de *Pf*tRip entière.

Après purification, la protéine est conservée à -20°C. Afin de vérifier si cette étape de congélation affecte l'activité de *Pf*tRip, les capacités à lier les ARNt des protéines congelées et fraîches ont été comparées par une expérience de retard sur gel (Figure 23).

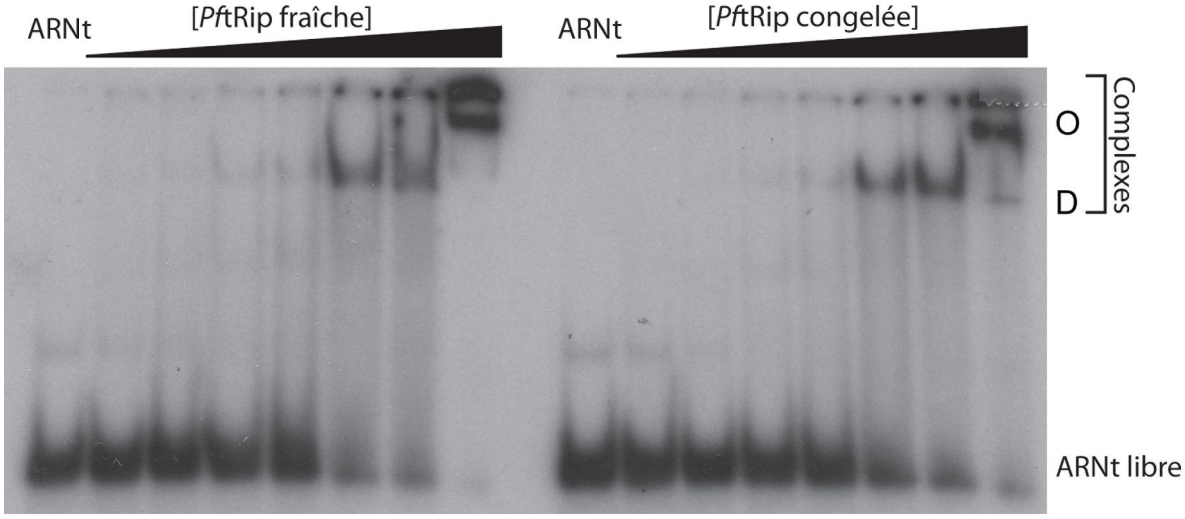



**Figure 13 : Effet de la congélation sur l'activité de *Pf*tRip**. La quantité d'ARNt marqué radioactivement est fixe et limitante, alors que la concentration de *Pf*tRip augmente (7 nM, 15 nM, 31 nM, 62 nM, 125 nM, 250 nM et 500 nM). Deux complexes sont visibles, ils correspondent à un ARNt lié à un dimère de *Pf*tRip (D) et à un ARNt lié à un oligomère de taille inconnu (O). Aux concentrations les plus élevées de *Pf*tRip, on observe des aggrégats qui restent bloqués dans les poches du gel.

La protéine est active et se comporte de la même manière avant et après la congélation. Cependant, dans les deux cas on constate que deux types de complexes sont formés (D et O). La présence de ces deux complexes a été observée précédemment et interprétée comme la coexistence d'un complexe correspondant à 1 ARNt lié à un dimère de

*Pf*tRip et d'un complexe où 1 ARNt est lié à deux dimères de *Pf*tRip (Bour, *et al*., 2016). Cependant, des études plus récentes menées dans le laboratoire par le Dr Anne Théobald-Dietrich, confirme l'existence d'un complexe primaire (D : un dimère de *Pf*tRip avec un ARNt), et ont mis à jour la capacité de *Pf*tRip à oligomériser (O). Cette oligomérisation est ARNt-dépendante, elle se déclenche quand l'ARNt est limitant, mais sa stoechiométrie n'est pas encore établie. Des expériences préliminaires de cryo-électromicroscopie indiquent cependant que cette stoechiométrie pourrait atteindre 6 dimères de *Pf*tRip pour 1 seul ARNt, suggérant qu'*in vivo*, au contact des ARNt exogènes, *Pf*tRip pourrait former un pore dans la membrane plasmique du parasite. Cependant, *in vitro* lorsque l'ARNt est en excès on observe uniquement le complexe D. C'est en se basant sur cette observation que les conditions expérimentales du MIST ont été définies (voir paragraphe 1.4.) : L'ARNt est toujours en excès par rapport à *Pf*tRip, d'une part pour former uniquement le complexe D et d'autre part pour permettre à *Pf*tRip de « choisir » les meilleurs interactants.

### *1.3. Purification de l'ARNt total de cellules HeLa*

La production d'ARNt total humain commence par l'extraction de l'ARN total de cellules HeLa (environ $50.10^6$ cellules). L'ARN est alors injecté en deux fois sur une colonne Superdex 200 et les ARNt sont séparés des autres ARN par exclusion de taille (Figure 24A). Sur le chromatogramme, le premier pic correspond aux grands ARN, tels que les ARN messagers ou les ARN ribosomaux alors que le deuxième pic, qui atteint une densité optique de 0,15 contient essentiellement des ARNt. La pureté de l'ARNt a été vérifiée sur un gel d'électrophorèse dénaturant (Figure 24B). La présence d'un troisième pic qui n'apparaît qu'au cours de la deuxième chromatographie correspond à des nucléotides et indique donc que les ARN sont partiellement dégradés lorsqu'ils sont stockés dans la boucle d'injection entre les deux purifications.

La quantité totale d'ARNt obtenu à partir de 50 millions de cellules HeLa est comprise entre 30 et 40 µg.

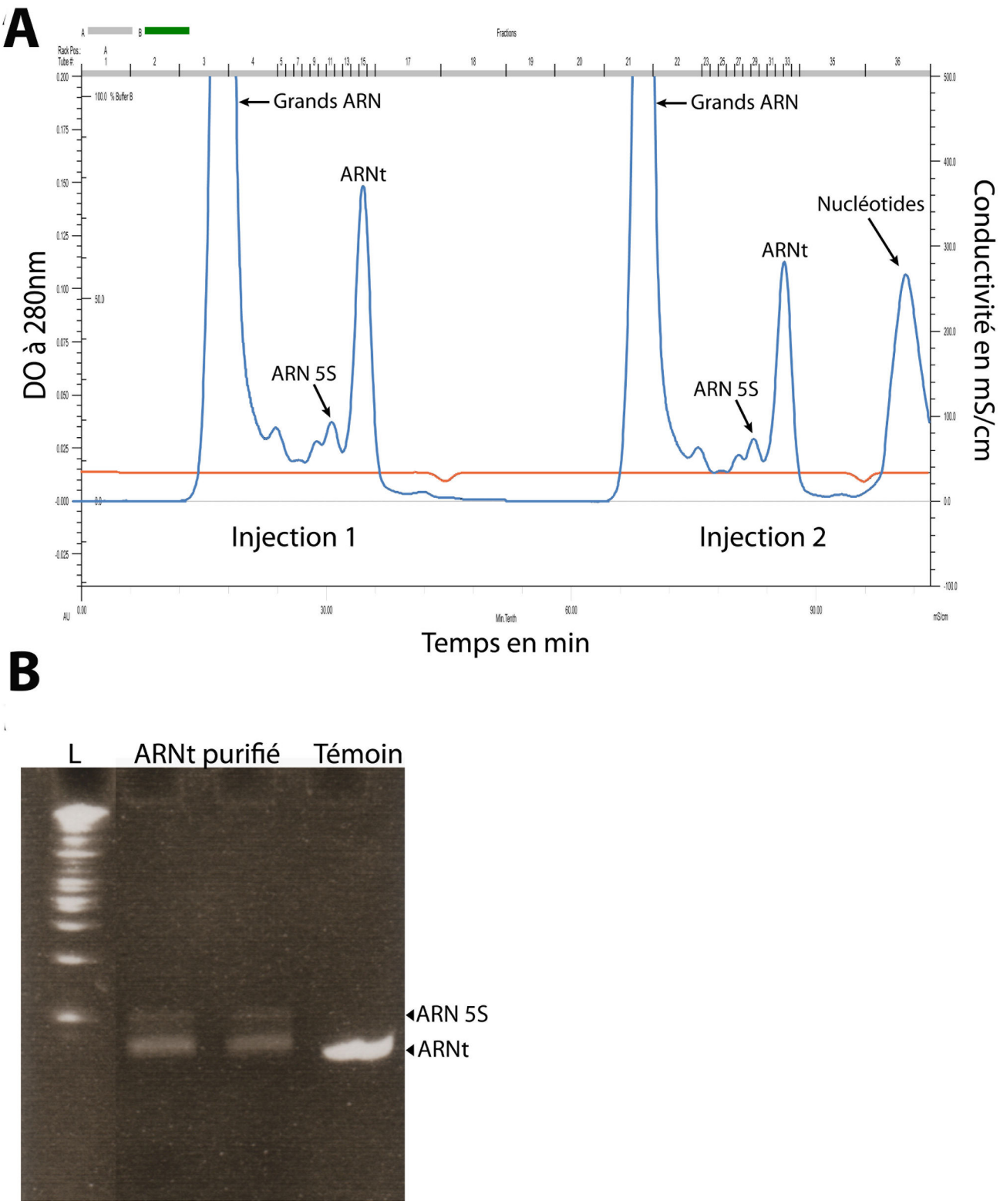



**Figure 24 : Purification des ARNt totaux extraits de cellules HeLa. (A)** profil d'élution des différents ARN sur une colonne d'exclusion de taille (Superdex 200). Le suivi de la conductivité est indiqué en rouge et la densité optique en bleu. Cette chromatographie permet de séparer efficacement les grands ARN des petits. **(B)** Le deuxième pic contient essentiellement des ARNt comme montré sur ce gel dénaturant 12%.

### *1.4. MIST*

*a. Isolement des complexes ARNt/PftRip*

Différents rapports ARNt/*Pf*tRip ont été utilisés pour former les complexes étudiés. Tandis que la protéine reste constante, la quantité d'ARNt a été augmentée dans chaque expérience pour avoir des rapports croissants entre ARNt et *Pf*tRip : 1 ARNt pour 1 dimère de *Pf*tRip, 4 ARNt pour 1 dimère, 10 pour 1 et 20 pour 1. Nous avons choisi de mettre un excès d'ARNt par rapport à la protéine pour augmenter la stringence au cours de l'étape de formation du complexe et permettre ainsi à *Pf*tRip de se fixer aux ARNt pour lesquels elle a le plus d'affinité. De plus dans ces conditions seul le complexe D contenant 1 ARNt par dimère de *Pf*tRip se forme.

La Figure 25 montre le résultat d'un gel natif où les complexes ARNt/*Pf*tRip sont séparés de l'ARNt libre. Dans la première condition (1 ARNt pour 1 dimère de *Pf*tRip), la faible concentration en ARNt conduit au schéma suivant : la bande correspondant au complexe D est presque invisible, alors que la bande correspondant au complexe O est bien présente plus haut dans le gel.

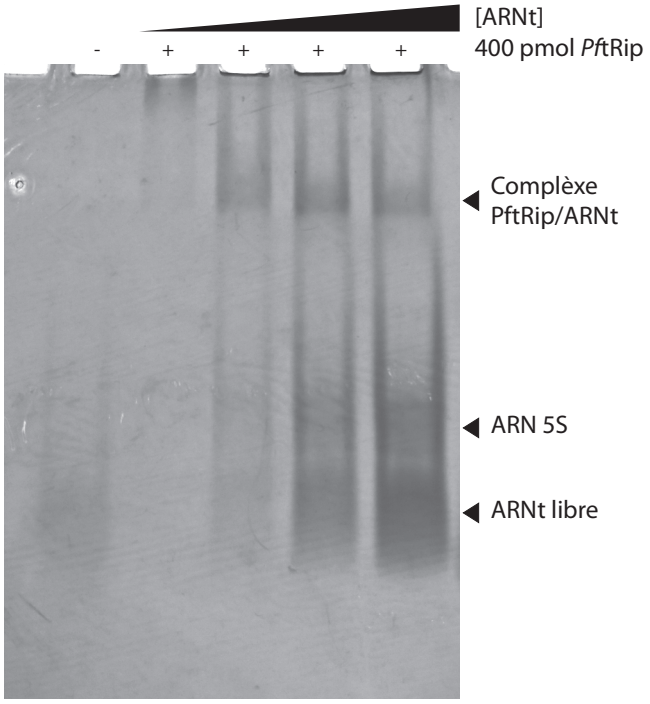



**Figure 25 : Etape de sélection des ARNt liés à *Pf*tRip.** Les complexes ARNt/*Pf*tRip sont séparés des ARNt libres sur gel natif 6 % et visualisés sous UV. La concentration de de dimères de *Pf*tRip est constante (200 nM), alors que la concentration d'ARNt total augmente (200, 800, 2000 et 4000 nM).

Dans les trois autres conditions testées (4, 10 et 20 ARNt par dimère de *Pf*tRip), on observe une bande unique d'intensité constante correspondant au complexe D, et une bande d'intensité croissante d'ARNt libres.

Un même échantillon d'ARNt total a été utilisé pour séparer les ARNt liés à *Pf*tRip des ARNt libres dans trois conditions expérimentales différentes (4, 10 et 20 ARNt par dimère de *Pf*tRip) au cours de deux expériences indépendantes (1 et 2). Les ARNt présents dans les bandes de gels contenant les complexes des expériences 1 et 2 et les fractions libres de l'expérience 1 ont été élués et analysés sur les puces à ARNt. En parallèle, un autre échantillon d'ARNt total purifié à partir de cellules HeLa a été analysé deux fois (Total 1 et 2) pour obtenir des puces à ARNt contrôles.

*b. Les puces à ARNt*

Chaque puce est composée de 48 sondes différentes (Figure Suppémentaire 1), répétées huit fois chacune. Elles s'hybrident soit à un isodécodeur unique (Arg-3h, Asn-1h, Asp-1h, Gly-2h, His-1h, Leu-3h, Lys-1h, Lys-3h, Lys4H, Met-1h, Sec-1h, Thr-3h, Thr-4h, Val-4h), soit à un seul isoaccepteur (Arg-1h, Arg-4h, Glu-2h, Ile-2h, Leu-2hn, Leu-4h, Leu-6h, Lys-2h, Met-2h, Phe-1h, Ser-1h, Ser-3h, Thr-1h, Thr-2h, Thr-5h, Tyr-2h), soit à plusieurs isoaccepteurs (Ala-2h, Arg-2h, Glu-1h, Gly-1h, Leu-1h, Pro-1h, Ser-2h, Val-1h), soit à deux ARNt différents (Glu-3h, Ser-4h, Thr6h, Tyr-1h, Val3H).

La Figure 26A montre le résultat obtenu pour une hybridation contrôle qui permet de quantifier la proportion de chaque ARNt dans l'échantillon initial (ARNt total). Des variations significatives entre les intensités de radioactivité sont observables pour chaque sonde, indiquant que, comme attendu, chaque ARNt est différemment représenté dans l'ARNt total humain (Dittmar, *et al*., 2006). La Figure 26B montre la même puce après hybridation des ARNt sélectionnés par *Pf*tRip dans le rapport 10 ARNt/dimère de *Pf*tRip. Pour chaque position, les intensités varient considérablement par rapport à l'hybridation témoin (Figure 26A), indiquant que certains ARNt sont enrichis dans le complexe avec *Pf*tRip par rapport à d'autres. Par exemple, l'ARNt$^{Arg}_{TCT}$ qui s'hybride à la sonde Arg-4h est peu abondant dans l'ARNt total (Figure 26A, cercles violets), mais est détecté plus fortement parmi les ARNt sélectionnés par *Pf*tRip (Figure 26B, cercles violets). On observe le phénomène inverse pour l'ARNt$^{Asp}_{GTC}$ reconnu par la sonde Asp-1h (cercles bleus). Ainsi *Pf*tRip fixe plus efficacement l'ARNt$^{Arg}_{TCT}$ que l'ARNt$^{Asp}_{GTC}$.

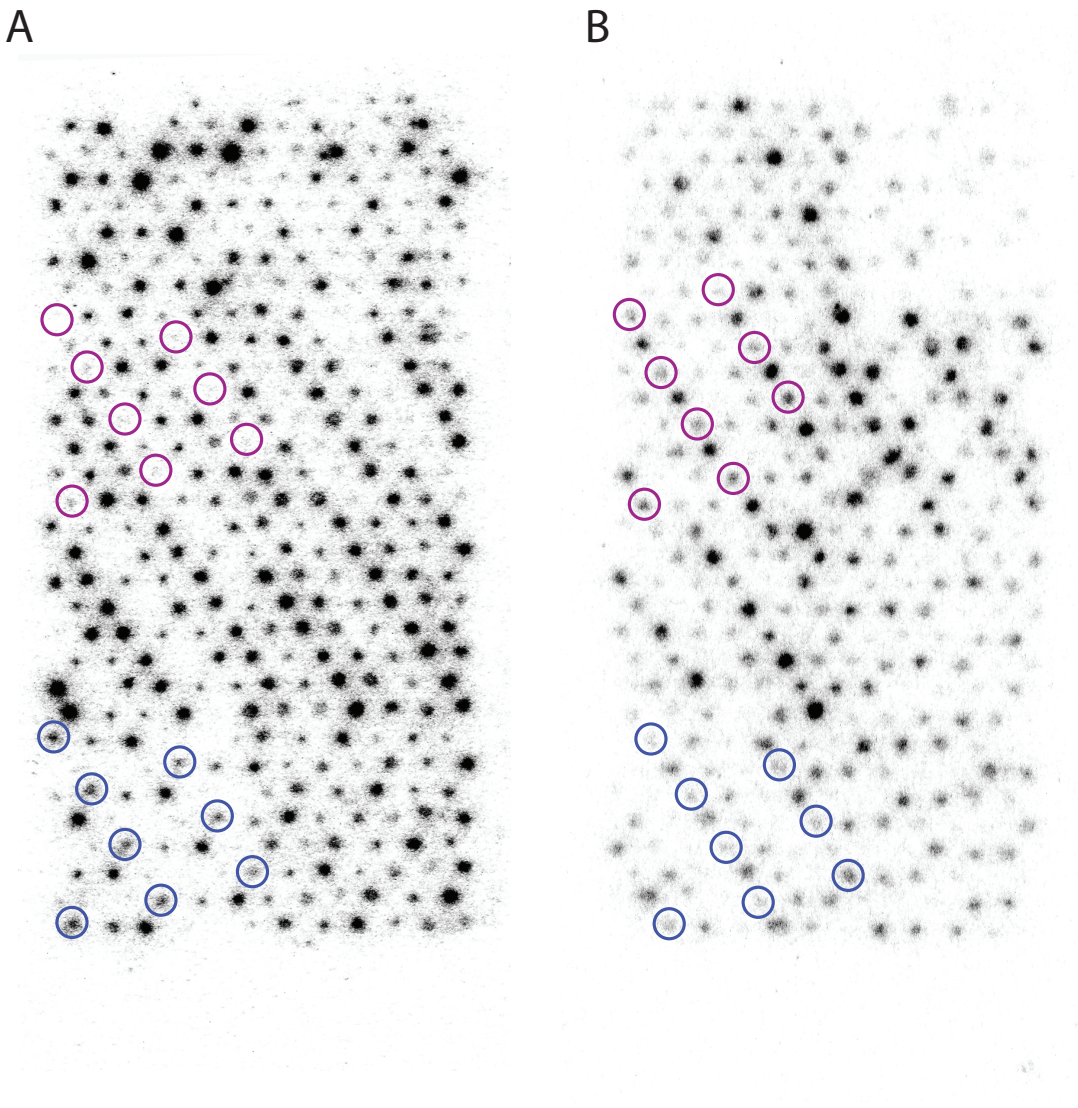



**Figure 26 : Exemples de puces à ARNt. (A)** Puce à ARNt contrôle hybridée à l'ARNt total de cellules HeLa (Total 1), marqué radioactivement. **(B)** Puce à ARNt hybridée à l'ARNt présent dans le complexe 1 (10 ARNt/dimère de *Pf*tRip). Les cercles violets indiquent la position des 8 copies de la sonde Arg-4h, et les cercles bleus, les 8 copies de la sonde Asp-1h. l'ARNt$^{Asp}_{GTC}$ reconnu par la sonde Asp-1h. Ainsi *Pf*tRip fixe plus efficacement l'ARNt$^{Arg}_{TCT}$ que l'ARNt$^{Asp}_{GTC}$.

*c. Quantification et interprétation des résultats*

Chaque puce a été quantifiée avec le programme ImageJ, et les moyennes des intensités correspondant aux 8 répétitions de chaque sonde ont été calculées. Les résultats bruts sont représentés dans le Tableau 2. Chaque cellule correspond à la proportion (en $^{0}/_{00}$) d'ARNt détecté spécifiquement par une sonde. Plus la quantité d'ARNt est élevée dans l'échantillon testé, plus la cellule est orange. A l'inverse, moins il est représenté, plus la

cellule est bleue. Certains ARNt ne sont pas détectables, ils correspondent aux sondes Ile-1h, Cys-1h, Gln-1h et Trp-1h, indiquant que ces sondes ne sont pas fonctionnelles.

Dans un premier test (Tableau 2A) deux contrôles ont été faits en hybridant de l'ARNt total de cellules HeLa sur deux puces différentes (Total 1 et Total 2). Les résultats sont comparables, confirmant que l'expérience est reproductible. On peut cependant observer des variations significatives d'une expérience à l'autre, en particuliers pour les ARNt qui sont peu abondants dans les échantillons testés. C'est le cas pour les ARNt détectés par les sondes Arg-2h, Leu-3h, Asn1h et Ser-2h où leurs proportions peuvent varier d'un facteur 4 d'une expérience à l'autre. Ces données n'ont pas été utilisées pour traiter les données des expériences de sélection par *Pf*tRip, car les ARNt totaux utilisés lors des expériences de sélection proviennent d'une purification différente. Cependant, la moyenne M(T) des deux expériences Total 1 et Total 2 a été utilisée pour ordonner l'ensemble des données expérimentales dans le Tableau 2B.

Les résultats bruts correspondant aux quantifications des puces à ARNt hybridées avec les ARNt provenant des complexes 1, 2 et ARNt libres 1 sont indiqués dans le Tableau 2B. Les données Complexe 1 et Complexe 2 sont identiques et montrent directement l'enrichissement ou l'appauvrissement de certains ARNt après la sélection. Les ARNt détectés par les sondes Leu-1h, Leu-3h, Asn-1h, Ser-2h et Val-4h sont significativement enrichis alors que les ARNt détectés par les sondes Arg-1h, Glu-1h, Ser-3h, Lys-2h et Tyr-2h sont sous représentés par rapport à leur concentration initiale dans l'échantillon.

De plus, on observe que les ARNt correspondants aux sondes Asp-1h, Thr-5h, Arg-1h, Arg-4h, Ile-21h et Val-3h se répartissent de façon différentielle entre les fractions « liés » et les fractions « libres ». C'est à dire que quand les ARNt liés sont abondants, les ARNt libres sont faibles, et réciproquement.

Enfin, la moyenne des sommes ARNt « liés + libres » de l'expérience 1 (M(+),2B) montre globalement le même profil que M(T) (Tableau 2A), les différences mineures observables étant certainement la conséquence de l'utilisation de deux lots d'ARNt total de cellules HeLa différents.

Les résultats définitifs sont représentés sous la forme « ARNt liés sur la somme des ARNt liés et libres » (Tableau 3). Ainsi les données traitées permettent de mettre

## Tableau 2 : Quantifications des puces à ARNt (données brutes).

**A**

| | Sans sélection | | |
|---|---|---|---|
| | Total 1 | Total 2 | M(T) |
| Ala-1h | 0,0 | 0,0 | . |
| Ile-1h | 0,0 | 0,0 | . |
| Cys-1h | 0,0 | 0,0 | . |
| Gln-1h | 0,0 | 0,0 | . |
| Trp-1h | 0,0 | 0,0 | . |
| Pro-1h | 8,5 | 2,2 | . |
| Arg-2h | 7,2 | 6,0 | . |
| Lys-3h | 2,8 | 11,3 | . |
| Leu-1h | 8,5 | 6,0 | . |
| Ser-4h | 7,7 | 7,4 | . |
| Arg-3h | 6,5 | 8,7 | . |
| Leu-3h | 6,4 | 10,1 | . |
| Asn-1h | 13,7 | 3,3 | . |
| Ser-2h | 3,8 | 13,3 | . |
| Asp-1h | 12,5 | 5,3 | . |
| Tyr-1h | 9,4 | 9,9 | . |
| Ser-1h | 6,2 | 14,1 | . |
| Sec-1h | 17,5 | 4,7 | . |
| Thr-2h | 9,2 | 14,3 | . |
| Val-4h | 16,7 | 10,9 | . |
| Thr-1h | 13,6 | 17,7 | . |
| Thr-5h | 11,4 | 21,5 | . |
| Glu-3h | 17,1 | 20,3 | . |
| His-1h | 22,6 | 15,4 | . |
| Val-1h | 16,1 | 26,5 | . |
| Gly-2h | 26,0 | 17,2 | . |
| Arg-1h | 27,0 | 16,4 | . |
| Gly-1h | 24,5 | 19,0 | . |
| Glu-1h | 25,6 | 19,2 | . |
| Ala-2h | 17,1 | 33,5 | . |
| Leu-2hn | 29,2 | 22,5 | . |
| Ser-3h | 17,3 | 35,3 | . |
| Lys-2h | 17,6 | 35,3 | . |
| Tyr-2h | 24,0 | 29,0 | . |
| Leu-4h | 26,8 | 34,5 | . |
| Thr6h | 23,6 | 39,2 | . |
| Lys-1h | 40,3 | 25,9 | . |
| Leu-6h | 32,5 | 37,9 | . |
| Lys-4h | 37,3 | 33,8 | . |
| Met-1h | 35,9 | 38,7 | . |
| Thr-3h | 45,0 | 29,7 | . |
| Thr-4h | 49,3 | 30,6 | . |
| Arg-4h | 33,9 | 47,8 | . |
| Ile-2h | 43,7 | 43,1 | . |
| Val-3h | 52,4 | 34,9 | . |
| Met-2h | 50,3 | 41,5 | . |
| Phe-1h | 48,4 | 47,7 | . |
| Glu-2h | 54,7 | 58,6 | . |

**B**

| | ARNt liés | | | | | | ARNt libres | | | liés + libres | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Complexe 1 | | | Complexe 2 | | | Libre 1 | | | 1 | | | |
| ARNt/tRip | 4 | 10 | 20 | 4 | 10 | 20 | 4 | 10 | 20 | 4 | 10 | 20 | M |
| Ala-1h | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 25,6 | 13,9 | 11,2 | 25,6 | 13,9 | 11,2 | |
| Ile-1h | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,7 | 0,0 | 0,0 | 0,7 | 0,0 | 0,0 | |
| Cys-1h | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | |
| Gln-1h | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | |
| Trp-1h | 0,0 | 3,4 | 0,0 | 1,6 | 0,0 | 0,0 | 0,0 | 0,9 | 0,0 | 0,0 | 4,3 | 0,0 | |
| Pro-1h | 3,7 | 3,5 | 2,2 | 2,1 | 1,8 | 7,0 | 19,7 | 18,7 | 22,0 | 23,4 | 22,3 | 24,3 | |
| Arg-2h | 5,4 | 7,9 | 1,1 | 15,4 | 0,0 | 18,2 | 6,7 | 8,5 | 6,7 | 12,0 | 16,4 | 7,8 | |
| Lys-3h | 3,0 | 6,9 | 7,1 | 7,8 | 8,5 | 1,1 | 13,1 | 7,2 | 1,3 | 16,1 | 14,0 | 8,4 | |
| Leu-1h | 4,9 | 33,5 | 21,0 | 17,3 | 24,8 | 19,2 | 0,0 | 0,5 | 0,0 | 4,9 | 34,1 | 21,0 | |
| Ser-4h | 11,6 | 22,9 | 13,4 | 9,7 | 10,5 | 10,7 | 0,0 | 2,7 | 0,0 | 11,6 | 25,6 | 13,4 | |
| Arg-3h | 8,3 | 4,0 | 7,0 | 3,3 | 8,0 | 3,4 | 0,6 | 6,6 | 5,7 | 8,9 | 10,6 | 12,8 | |
| Leu-3h | 18,2 | 14,7 | 21,5 | 1,6 | 21,3 | 5,2 | 11,2 | 21,1 | 18,4 | 29,4 | 35,8 | 39,9 | |
| Asn-1h | 14,0 | 18,4 | 21,3 | 13,8 | 10,1 | 21,1 | 0,0 | 1,9 | 5,9 | 14,0 | 20,3 | 27,2 | |
| Ser-2h | 6,6 | 21,5 | 21,6 | 15,2 | 21,5 | 3,3 | 4,1 | 9,0 | 2,2 | 10,8 | 30,5 | 23,7 | |
| Asp-1h | 4,8 | 2,0 | 0,5 | 2,4 | 0,0 | 1,1 | 25,7 | 21,7 | 19,8 | 30,5 | 23,7 | 20,3 | |
| Tyr-1h | 1,4 | 2,6 | 1,6 | 4,1 | 0,7 | 1,3 | 13,2 | 8,8 | 2,3 | 14,6 | 11,4 | 3,9 | |
| Ser-1h | 0,0 | 0,0 | 0,0 | 0,0 | 3,6 | 0,0 | 15,0 | 13,8 | 8,7 | 15,0 | 13,8 | 8,7 | |
| Sec-1h | 5,8 | 9,9 | 6,4 | 2,1 | 1,8 | 6,2 | 32,6 | 27,9 | 31,9 | 38,4 | 37,8 | 38,4 | |
| Thr-2h | 9,4 | 0,0 | 2,5 | 2,8 | 2,4 | 0,0 | 18,2 | 5,5 | 20,1 | 27,6 | 5,5 | 22,6 | |
| Val-4h | 14,1 | 17,3 | 17,8 | 16,4 | 14,8 | 24,4 | 0,7 | 4,7 | 8,5 | 14,7 | 22,0 | 26,3 | |
| Thr-1h | 14,9 | 5,0 | 11,5 | 0,3 | 6,2 | 0,0 | 11,4 | 13,0 | 15,6 | 26,3 | 18,0 | 27,0 | |
| Thr-5h | 0,0 | 0,0 | 0,9 | 2,6 | 8,6 | 0,0 | 64,1 | 39,5 | 32,1 | 64,1 | 39,5 | 33,0 | |
| Glu-3h | 18,9 | 13,8 | 18,5 | 7,8 | 13,8 | 12,2 | 6,3 | 18,1 | 15,6 | 25,2 | 31,9 | 34,0 | |
| His-1h | 38,4 | 35,3 | 28,0 | 21,9 | 14,6 | 38,2 | 21,9 | 25,5 | 26,2 | 60,3 | 60,8 | 54,1 | |
| Val-1h | 21,5 | 8,3 | 12,3 | 5,3 | 12,1 | 9,3 | 7,7 | 11,6 | 16,5 | 29,1 | 19,9 | 28,8 | |
| Gly-2h | 29,1 | 21,5 | 23,3 | 18,3 | 17,0 | 28,5 | 22,0 | 26,5 | 36,4 | 51,2 | 48,0 | 59,7 | |
| Arg-1h | 8,3 | 10,6 | 9,0 | 11,2 | 8,5 | 3,4 | 91,1 | 61,1 | 48,9 | 99,4 | 71,7 | 57,9 | |
| Gly-1h | 43,6 | 27,4 | 26,9 | 19,6 | 16,8 | 15,1 | 20,8 | 31,3 | 30,5 | 64,5 | 58,7 | 57,4 | |
| Glu-1h | 11,2 | 14,1 | 9,4 | 23,3 | 11,0 | 7,3 | 50,3 | 42,8 | 28,3 | 61,5 | 56,9 | 37,7 | |
| Ala-2h | 25,1 | 38,1 | 47,4 | 45,5 | 67,3 | 34,6 | 0,0 | 7,9 | 9,3 | 25,1 | 45,9 | 56,7 | |
| Leu-2hn | 29,3 | 49,1 | 35,4 | 38,2 | 25,8 | 60,3 | 5,7 | 14,6 | 10,9 | 35,0 | 63,7 | 46,3 | |
| Ser-3h | 11,3 | 14,7 | 18,8 | 30,5 | 27,7 | 5,4 | 9,6 | 7,0 | 11,9 | 21,0 | 21,7 | 30,7 | |
| Lys-2h | 18,0 | 7,7 | 14,0 | 7,6 | 18,5 | 3,8 | 50,7 | 21,3 | 26,5 | 68,7 | 29,0 | 40,4 | |
| Tyr-2h | 24,0 | 10,5 | 16,3 | 9,1 | 15,1 | 8,6 | 15,5 | 16,3 | 19,8 | 39,5 | 26,8 | 36,1 | |
| Leu-4h | 30,4 | 28,5 | 25,5 | 28,7 | 28,3 | 33,1 | 34,3 | 31,8 | 32,2 | 64,7 | 60,2 | 57,7 | |
| Thr6h | 23,6 | 49,6 | 38,6 | 60,3 | 50,3 | 19,8 | 15,6 | 34,6 | 31,0 | 39,2 | 84,2 | 69,6 | |
| Lys-1h | 35,5 | 39,1 | 30,5 | 32,5 | 24,8 | 46,8 | 46,4 | 40,9 | 48,8 | 81,9 | 80,0 | 79,3 | |
| Leu-6h | 33,7 | 31,4 | 21,1 | 46,5 | 37,0 | 60,2 | 48,8 | 45,7 | 47,0 | 82,4 | 77,1 | 68,1 | |
| Lys-4h | 27,8 | 36,5 | 40,7 | 39,1 | 43,0 | 37,8 | 27,7 | 29,1 | 20,7 | 55,4 | 65,5 | 61,4 | |
| Met-1h | 41,9 | 54,6 | 38,6 | 34,2 | 31,7 | 53,5 | 31,1 | 36,8 | 29,6 | 73,0 | 91,5 | 68,2 | |
| Thr-3h | 45,5 | 37,3 | 36,6 | 45,9 | 33,9 | 46,4 | 17,0 | 41,2 | 48,3 | 62,5 | 78,4 | 84,9 | |
| Thr-4h | 48,8 | 41,4 | 42,7 | 49,3 | 39,9 | 60,0 | 28,2 | 42,0 | 49,3 | 77,0 | 83,4 | 91,9 | |
| Arg-4h | 28,2 | 17,7 | 27,7 | 21,7 | 30,8 | 31,7 | 8,0 | 12,7 | 19,3 | 36,2 | 30,4 | 47,0 | |
| Ile-2h | 28,4 | 37,4 | 56,3 | 50,2 | 61,1 | 24,8 | 19,3 | 14,4 | 12,4 | 47,7 | 51,8 | 68,7 | |
| Val-3h | 71,2 | 51,2 | 64,7 | 81,6 | 74,5 | 92,9 | 4,8 | 22,5 | 36,1 | 76,1 | 73,8 | 100,8 | |
| Met-2h | 58,2 | 49,7 | 59,1 | 35,8 | 52,4 | 25,6 | 38,5 | 36,0 | 55,7 | 96,7 | 85,7 | 114,8 | |
| Phe-1h | 77,7 | 44,4 | 58,6 | 54,1 | 59,9 | 53,6 | 29,4 | 32,4 | 35,7 | 107,1 | 76,9 | 94,3 | |
| Glu-2h | 44,4 | 56,5 | 42,8 | 63,2 | 40,0 | 65,1 | 86,7 | 70,0 | 40,9 | 131,1 | 126,5 | 83,8 | |

Les ARNt présents dans deux échantillons d'ARNt total (Total 1 et Total 2), dans ou en dehors des complexes avec *Pf*tRip (Complexe 1, Complexe 2, libre 1) ont été hybridés aux puces à ARNt et quantifiés par ImageJ. Les données sont indiquées en $^{0}/_{00}$. Les chiffres 4, 10, 20 correspondent aux conditions de formation du complexe, 4, 10 ou 20 ARNt par dimère de *Pf*tRip, respectivement. La couleur orange correspond aux ARNt les plus représentés dans les échantillons et le bleu correspond aux ARNt les moins représentés.

directement en évidence l'accumulation des ARNt caractérisés par une affinité élevée (orange) et au contraire la diminution des ARNt avec une faible affinité pour tRip (bleu).

Ces résultats montrent clairement que *Pf*tRip ne reconnaît pas tous les ARNt de la même manière. Pour chaque ARNt, l'efficacité de fixation reste globalement constante quel que soit le rapport ARNt/*Pf*tRip utilisé. De plus l'affinité des ARNt n'est pas dépendante de leur concentration dans l'échantillon initial, indiquant que la sélection a été efficace (Tableau 3A). En se basant sur la moyenne des trois valeurs liés/liés+libres du complexe 1 (M(Co), Tableau 3B), il existe un facteur 25 entre le ligant le plus faible (l'ARNt$^{Ser}_{CGA}$/sonde Ser-1h) et le ligant le plus fort (ARNt$^{Leu}_{wAG}$/Leu1h). Les deux sondes Met-1h et Thr-6h (bleu ciel), qui reconnaissent le même ARNt$^{Met}$i donnent exactement le même résultat, confirmant la bonne reproductibilité des expériences.

De façon évidente, la séquence des ARNt ne joue aucun rôle dans leur interaction avec *Pf*tRip. En effet, les différents isoaccepteurs d'un même ARNt peuvent avoir des affinités très différentes pour *Pf*tRip (Tableau 3B), comme c'est le cas pour l'ARNt$^{Ser}_{CGA}$ et l'ARNt$^{Ser}_{wGA}$ (en rouge, sonde Ser-1h et Ser-2h, respectivement) ou l'ARNt$^{Ala}_{IGC}$ et l'ARNt$^{Ala}_{hGC}$ (sondes Ala-1h et Ala-2h, violet). Un isoaccepteur n'est pas du tout retenu par *Pf*tRip alors que l'autre est un des meilleurs ligands. On observe exactement le même comportement pour des isodécodeurs ; parmis les trois isodécodeurs de l'ARNt$^{Lys}_{TTT}$ testés en vert, l'ARNt$^{Lys}_{TTT}$-1 (sonde Lys-2h) est moins bien reconnu que les deux autres (sondes Lys-3h et lys-4h).

Les ARNt ayant les affinités les plus faibles pour *Pf*tRip sont les ARNt $Ala_{IGC}$,$Ser_{CGA}$, $Thr_{CGT}$, $Thr_{TGT}$, $Asp_{GTC}$, $Arg_{ICG}$, $Pro_{hGG}$, $Sec_{TCA}$, $Tyr_{GCA}$ et $Glu_{yTC}$ et les ARNt se liant le mieux à *Pf*tRip correspondent aux ARNt $Ile_{TAT}$, $Ser_{wGA/AGA}$, $Val_{TAC}$, $Asn_{GTT}$, $Ala_{hGC}$, $Ser_{wGA}$ et $Leu_{wAG/CAG}$.

### *1.5. Etude de l'affinité de* Pft*Rip pour différents transcrits d'ARNt* in vitro

L'étape suivante a été de confirmer les résultats obtenus par la technique du MIST. Les séquences des ARNt $Ala_{AGC}$ (Ala-1h), $Thr_{CGT}$-2 (Thr-5h), $Asp_{GTC}$ (Asp-1h), $Arg_{ACG}$ (Arg-1h) $Lys_{TTT}$-1 (Lys-2h), $Tyr_{GCA}$ (Tyr-1h), $Tyr_{GTA}$ (Tyr-2h), $Ser_{AGA}$ (Ser2h), $Asn_{GTT}$ (Asn-1h) et $Leu_{TAG}$ (Leu-1h) ont été clonées et transcrites *in vitro* (Figure Supplémentaire 2, Tableau 4).

La Figure 27 permet de visualiser l'ensemble des transcrits purifiés sur gel dénaturant. On observe que l'ARNt$^{Tyr}_{GCA}$ migre plus lentement que les autres transcrits de la même taille (75 nucléotides), impliquant son exclusion des tests de retard sur gel.

## Tableau 3 : Traitement des données de MIST

**A**

| M(T) | ARNt/tRip | liés 4 | liés 10 | liés 20 | liés+libres 4 | liés+libres 10 | liés+libres 20 | liés/(liés+libres) 4 | liés/(liés+libres) 10 | liés/(liés+libres) 20 | liés/(liés+libres) M(Co) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ala-1h | | | | | | | | | | |
| | Ile-1h | | | | | | | | | | |
| | Cys-1h | | | | | | | | | | |
| | Gln-1h | | | | | | | | | | |
| | Trp-1h | | | | | | | | | | |
| | Pro-1h | | | | | | | | | | |
| | Arg-2h | | | | | | | | | | |
| | Lys-3h | | | | | | | | | | |
| | Leu-1h | | | | | | | | | | |
| | Ser-4h | | | | | | | | | | |
| | Arg-3h | | | | | | | | | | |
| | Leu-3h | | | | | | | | | | |
| | Asn-1h | | | | | | | | | | |
| | Ser-2h | | | | | | | | | | |
| | Asp-1h | | | | | | | | | | |
| | Tyr-1h | | | | | | | | | | |
| | Ser-1h | | | | | | | | | | |
| | Sec-1h | | | | | | | | | | |
| | Thr-2h | | | | | | | | | | |
| | Val-4h | | | | | | | | | | |
| | Thr-1h | | | | | | | | | | |
| | Thr-5h | | | | | | | | | | |
| | Glu-3h | | | | | | | | | | |
| | His-1h | | | | | | | | | | |
| | Val-1h | | | | | | | | | | |
| | Gly-2h | | | | | | | | | | |
| | Arg-1h | | | | | | | | | | |
| | Gly-1h | | | | | | | | | | |
| | Glu-1h | | | | | | | | | | |
| | Ala-2h | | | | | | | | | | |
| | Leu-2hn | | | | | | | | | | |
| | Ser-3h | | | | | | | | | | |
| | Lys-2h | | | | | | | | | | |
| | Tyr-2h | | | | | | | | | | |
| | Leu-4h | | | | | | | | | | |
| | Thr6h | | | | | | | | | | |
| | Lys-1h | | | | | | | | | | |
| | Leu-6h | | | | | | | | | | |
| | Lys-4h | | | | | | | | | | |
| | Met-1h | | | | | | | | | | |
| | Thr-3h | | | | | | | | | | |
| | Thr-4h | | | | | | | | | | |
| | Arg-4h | | | | | | | | | | |
| | Ile-2h | | | | | | | | | | |
| | Val-3h | | | | | | | | | | |
| | Met-2h | | | | | | | | | | |
| | Phe-1h | | | | | | | | | | |
| | Glu-2h | | | | | | | | | | |

**B**

| Sonde | ARNt | M(T) | M(Co) |
|---|---|---|---|
| Ala-1h | AlaIGC | | 0 |
| Ser-1h | SerCGA | | 1 |
| Thr-5h | ThrCGT-2 | | 1 |
| Asp-1h | AspGTC | | 2 |
| Arg-1h | ArgiCG | | 3 |
| Pro-1h | ProhGG | | 4 |
| Sec-1h | SecTCA | | 4 |
| Thr-2h | ThrTGT-1 | | 4 |
| Tyr-1h | CysGCA | | 6 |
| Glu-1h | GluyTC | | 6 |
| Lys-2h | LysTTT-1 | | 7 |
| Thr-1h | ThrmGT | | 8 |
| Lys-1h | LysCTT | | 11 |
| Leu-6h | LeuTAA-2 | | 11 |
| Tyr-2h | TyrGTA | | 11 |
| Leu-3h | LeuCAA | | 12 |
| Gly-2h | GlyTCC | | 12 |
| Glu-2h | GluTTC-1 | | 12 |
| Leu-4h | LeuTAA-1 | | 12 |
| Gly-1h | GlysCC | | 12 |
| Val-1h | ValmAC | | 13 |
| Arg-2h | ArgyCG | | 13 |
| Lys-3h | LysTTT-2 | | 13 |
| Met-2h | Met-e | | 13 |
| Glu-3h | GluTTC-2 | | 14 |
| His-1h | HisGTG | | 14 |
| Thr-4h | ThrTGT-2 | | 14 |
| Thr-3h | ThrCGT-1 | | 14 |
| Met-1h | Met-i | | 15 |
| Lys-4h | LysTTT-3 | | 15 |
| Thr6h | Met-i | | 15 |
| Arg-3h | ArgCCT | | 16 |
| Phe-1h | PheGAA | | 16 |
| Ser-3h | SerGCT | | 17 |
| Arg-4h | ArgTCT | | 17 |
| Ile-2h | IleTAT | | 19 |
| Ser-2h | SerwGA | | 19 |
| Val-3h | AlaTGC | | 20 |
| Leu-2hn | LeuCAG | | 20 |
| Val-4h | ValTAC | | 21 |
| Asn-1h | AsnGTT | | 23 |
| Ala-2h | AlahGC | | 23 |
| Ser-4h | SerAGA | | 24 |
| Leu-1h | LeuwAG | | 25 |

r : A,G  s : C,G
y : C,T  w : A,T
m : A,C  h : A,C,T

liés = Moyenne (Complexe1/Complexe 2)

**Tableau 4 : Tableau récapitulatif des ARNt transcrits *in vitro***

| Sonde | transcrit | M(Co) |
|---|---|---|
| **Ala-1h** | AlaAGC | **0** |
| **Thr-5h** | ThrCGT-2 | **1** |
| **Asp-1h** | AspGTC | **2** |
| **Arg-1h** | ArgACG | **3** |
| **Lys-2h** | LysTTT-1 | **7** |
| **Tyr-1h** | TyrGCA | **6** |
| **Tyr-2h** | TyrGTA | **11** |
| **Ser-2h** | SerAGA | **19** |
| **Asn-1h** | AsnGTT | **23** |
| **Leu-1h** | LeuTAG | **25** |

La sonde correspondant à chaque transcrit, ainsi que leur affinité relative sont indiquées.

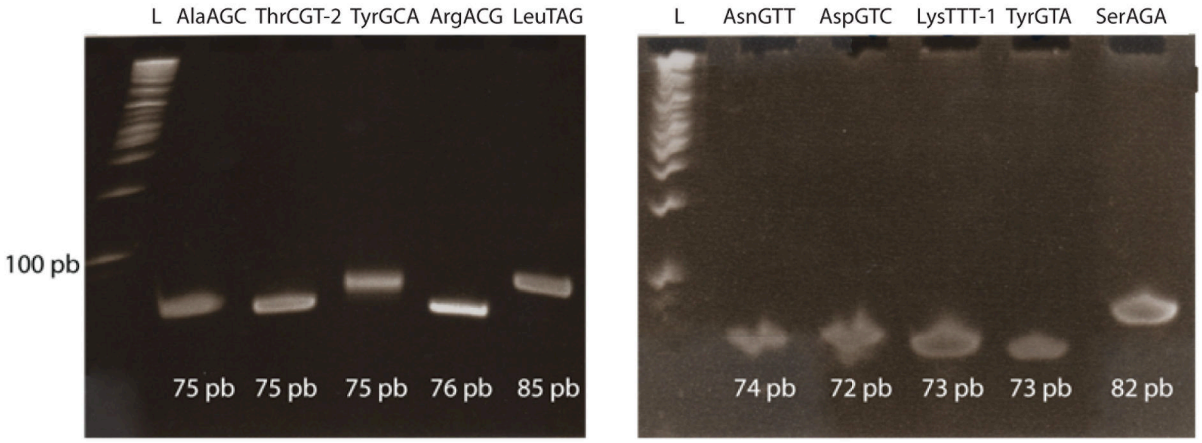



**Figure 27 : Qualité des transcrits d'ARNt sur gel dénaturant 12%.** L correspond à l'échelle de poids moléculaire.

Au moins trois expériences de retard sur gel ont été réalisées avec chaque transcrit (Figure 28). Pour cela, nous avons choisi de faire des compétitions : un complexe est formé entre *Pf*tRip et l'ARNt total de cellules HeLa marqué radioactivement. Des concentrations croissantes (de 25 nM à 800 nM) de transcrits froids compétiteurs sont ajoutés. En fonction de leur affinité pour *Pf*tRip, chaque transcrit a la capacité de déplacer l'ARNt total plus ou moins efficacement. En d'autres mots, pour une même concentration en transcrits compétiteurs dans le test, plus *Pf*tRip a d'affinité pour le transcrit, moins il y aura d'ARNt marqué dans le complexe. On observe sur tous les gels de retard de la Figure 28 que l'intensité de la bande correspondant au complexe *Pf*tRip/ARNt total diminue quand la

concentration en compétiteur augmente. Pour chaque concentration en compétiteur, les bandes ARNt lié et ARNt libre ont été quantifiées avec le programme ImageJ. Afin de comparer les efficacités de dissociation (E(diss)) de chaque transcrit, la formule suivante a été appliquée :

$$E(diss) = \left(\frac{ARNtlié}{ARNtlié + ARNtlibre}\right)_{50nM} - \left(\frac{ARNtlié}{ARNtlié + ARNtlibre}\right)_{100\ nM}$$

Cette formule permet de comparer directement les vitesses de dissociation du complexe *Pf*tRip/ARNt total en présence de chaque transcrit compétiteur entre les concentrations 50 nM et 100 nM (zone linéaire). Plus E(diss) est grand, plus le transcrit dissocie efficacement le complexe, et plus il a une affinité élevée pour *Pf*tRip. Pour chaque transcrit, les moyennes et les écart types ont été calculés (Figure 29). Les transcrits ne se comportent pas comme les ARNt correspondants dans les expériences de MIST. D'une part, les valeurs-p calculées pour les E(diss) ne sont pas significatives (entre 0,16 et 0,71). D'autre part, les E(diss) des transcrits ne corrèlent pas avec les affinités déterminées par MIST pour les ARNt correspondants (Figure 29A). Les moyennes des E(diss) calculées varient au maximum d'un facteur 3,25 entre 0,12 (pour $tr^{Tyr}_{GTA}$ et $tr^{Asn}_{GTT}$) et0,39 ($tr^{Ala}_{AGC}$) (Figure 29B), ce qui est faible par rapport au facteur 25 observé dans l'expérience de MIST (Figure 29A). Ces observations confirment que les séquences des ARNt ne jouent pas de rôle prépondérant dans la reconnaissance des ARNt par *Pf*tRip. En conséquence les transcrits d'ARNt ne permettent pas d'identifier les signaux moléculaires impliqués dans la reconnaissance des ARNt par *Pf*tRip.

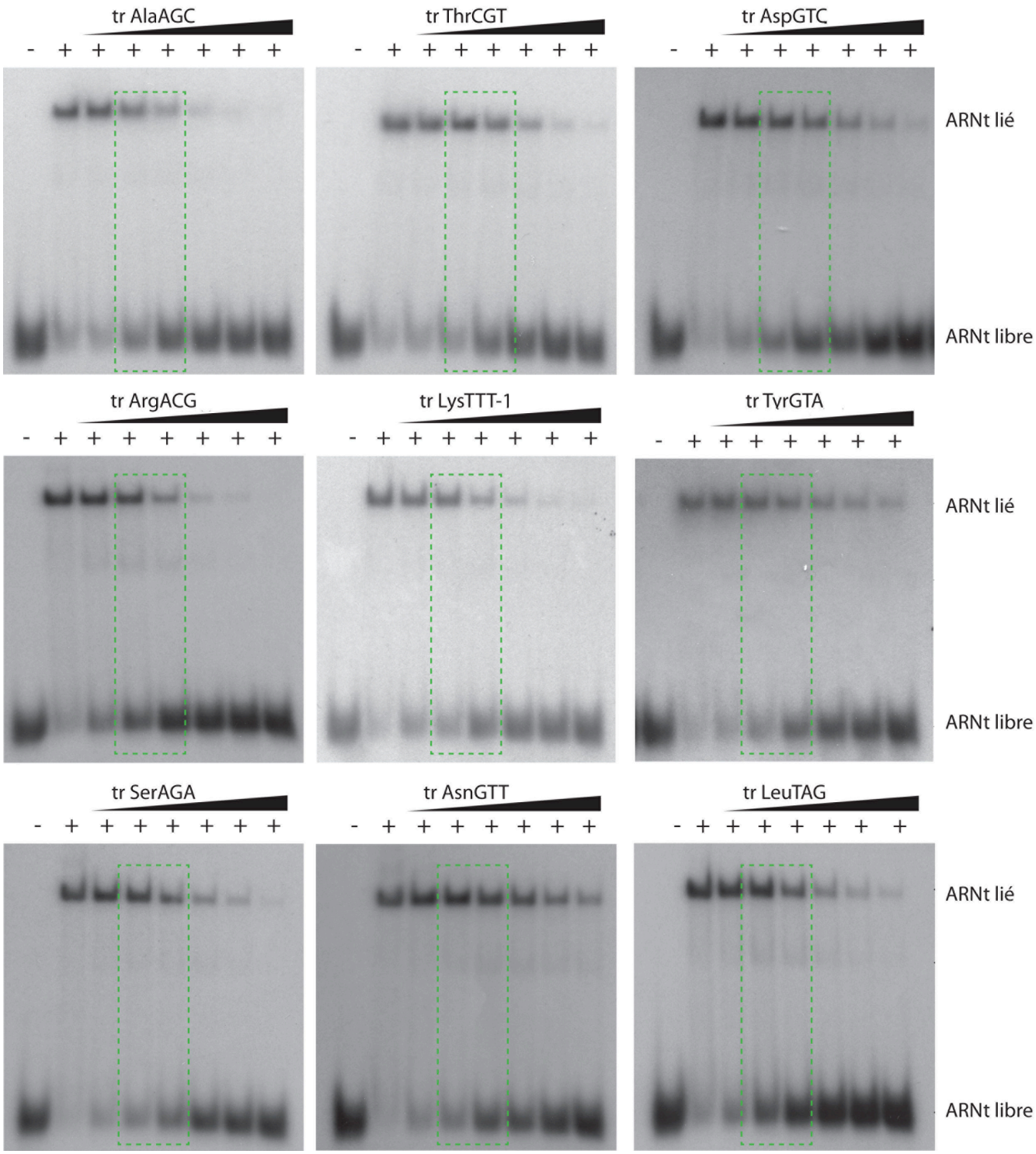



**Figure 28 : Un exemple d'expériences de retard sur gel pour les transcrits testés.** Sur chaque gel, le premier puits ne contient pas de protéine et sert de contrôle négatif, le deuxième puits contient uniquement l'ARNt total radioactif et *Pf*tRip, comme contrôle positif. Les 6 autres puits contiennent en plus des concentrations croissantes de transcrits compétiteurs qui varient de 25 à 800 nM). Les conditions 50 et 100 nM de compétiteurs, utilisées pour calculer l'efficacité de dissociation de chaque transcrit sont encadrées en vert.

**A**

| Sonde | transcrit | M(Co) | E(diss) | n |
|---|---|---|---|---|
| **Ala-1h** | AlaAGC | **0** | **0,39** | **3** |
| **Thr-5h** | ThrCGT-2 | **1** | **0,15** | **5** |
| **Asp-1h** | AspGTC | **2** | **0,14** | **3** |
| **Arg-1h** | ArgACG | **3** | **0,23** | **4** |
| **Lys-2h** | LysTTT-1 | **7** | **0,22** | **3** |
| **Tyr-2h** | TyrGTA | **11** | **0,12** | **4** |
| **Ser-2h** | SerAGA | **19** | **0,20** | **4** |
| **Asn-1h** | AsnGTT | **23** | **0,12** | **3** |
| **Leu-1h** | LeuTAG | **25** | **0,17** | **4** |

**B**

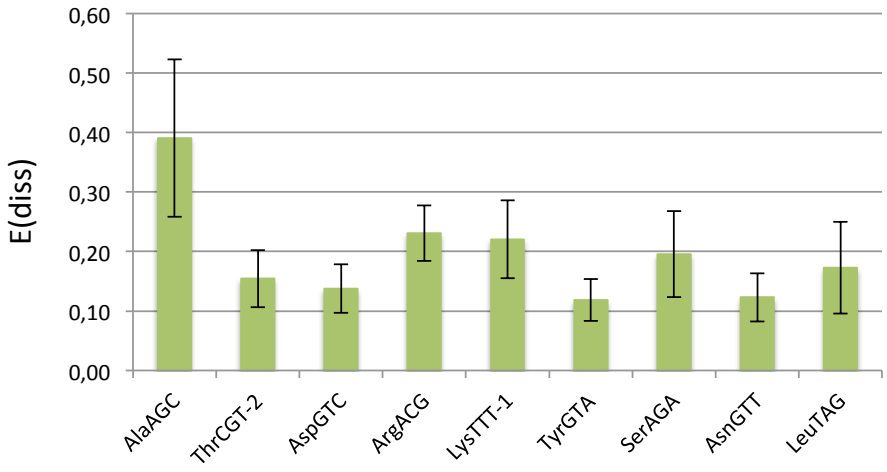



**Figure 29 : Détermination des affinités relatives des transcrits pour *Pf*tRip. (A)** Les affinités relatives ont été déterminées par la technique MIST pour les ARNt natifs et par retards sur gel pour les transcrits correspondants. Le code couleur indique les affinités les plus fortes en orange et les plus faibles en bleu. (n) correspond au nombre de fois où les expériences de retard sur gel ont été effectuées pour chaque transcrit. **(B)** Les moyennes des E(diss) et les écart type calculés sont représentés sous forme de graphe pour chaque transcrit.

### *1.6. Rôle des modifications post-transcriptionnelles dans la discrimination des ARNt par Pf*tRip

En dehors de leur séquence primaire, les ARNt sont caractérisés par la présence de nombreuses modifications post-transcriptionnelles. Celles-ci sont localisées en majorité dans le coude de la structure en L reconnu par *Pf*tRip. En particulier, ces modifications post-transcriptionnelles sont importantes pour le repliement correct de l'ARNt et sa stabilité

(Revu dans Giegé, 2008, Motorin et Helm, 2010, et Hopper, 2013). Afin de tester le rôle que jouent ces bases modifiées dans la reconnaissance des ARNt par *Pf*tRip, 4 ARNt natifs ont été choisis (Tableau 5) pour être purifiés à partir de cellules HeLa et effectuer des expériences de retard sur gel.

**Tableau 5 : Rendement des purifications des ARNt natifs à partir d'ARNt total de cellules HeLa.**

| Sonde | ARNt | M(Co) | % dans M(T) | théorique (µg) | obtenu (µg) | Rendement (%) |
|---|---|---|---|---|---|---|
| **Asp-1h** | AspGTC | **2** | 0,89 | 3,38 | 8 | 240 |
| **Arg-1h** | ArgiCG | **3** | 2,17 | 8,24 | 5 | 60 |
| **Asn-1h** | AsnGTT | **23** | 0,85 | 3,23 | 1,8 | 55 |
| **Ala-2h** | AlahGC | **23** | 2,53 | 9,61 | 1,8 | 19 |

Quatre ARNt ont été purifiés à partir d'ARNt total de cellules HeLa : deux ARNt pour lesquels *Pf*tRip a une bonne affinité ($ARNt^{Ala}_{hGC}$ et $ARNt^{Asn}_{GTT}$), et deux ARNt pour lesquels, la protéine a peu d'affinité selon les résultats du MIST ($ARNt^{Arg}_{ICG}$ et $ARNt^{Asp}_{GTC}$) (Tableau 5). Cette purification a été réalisée par hybridation de l'ARNt choisi à une sonde biotinylée, suivi d'une purification sur une résine de streptavidine. Les sondes biotinylées ont été conçues pour s'hybrider à la zone située entre la boucle D et la boucle de l'anticodon et en se basant sur les sondes conçues pour les puces à ARNt utilisés dans le MIST.

Les résultats obtenus avec les puces contrôles ont permis d'évaluer la quantité théorique de chaque ARNt dans les cellules HeLa (Tableau 2A). Pour chaque purification, 375 µg d'ARNt total extrait de cellules HeLa (correspond à l'ARNt total purifié à partir de 700 millions de cellules) ont été nécessaires pour obtenir les quantités d'ARNt purs indiquées dans le Tableau 5. Les rendements varient fortement entre les différents ARNt. En particulier, les 240 % d'enrichissement après purification de l'$ARNt^{Asp}_{GTC}$ suggèrent (i) que cet ARNt n'est pas pur ou (i) que notre préparation d'ARNt de cellules HeLa est fortement enrichie en $ARNt^{Asp}$.

Pour évaluer la pureté et identifier les modifications post-transcriptionnelles des ARNt natifs obtenus, 0,5 µg de chaque ARNt a été analysé par spectrométrie de masse. Ces analyses ont confirmé la pureté des $ARNt^{Asp}_{GTT}$ et $ARNt^{Arg}_{ICG}$ et ont permis de confirmer ou d'identifier les modifications post-transcriptionnelles présentes, respectivement (Figures

30A et 30B). Cependant, les analyses des échantillons correspondant à l'ARNt$^{Ala}_{hGC}$ et l'ARNt$^{Asn}_{GTT}$ identifient de nombreux contaminants, rendant ces échantillons impropres pour nos expériences de retard sur gel.

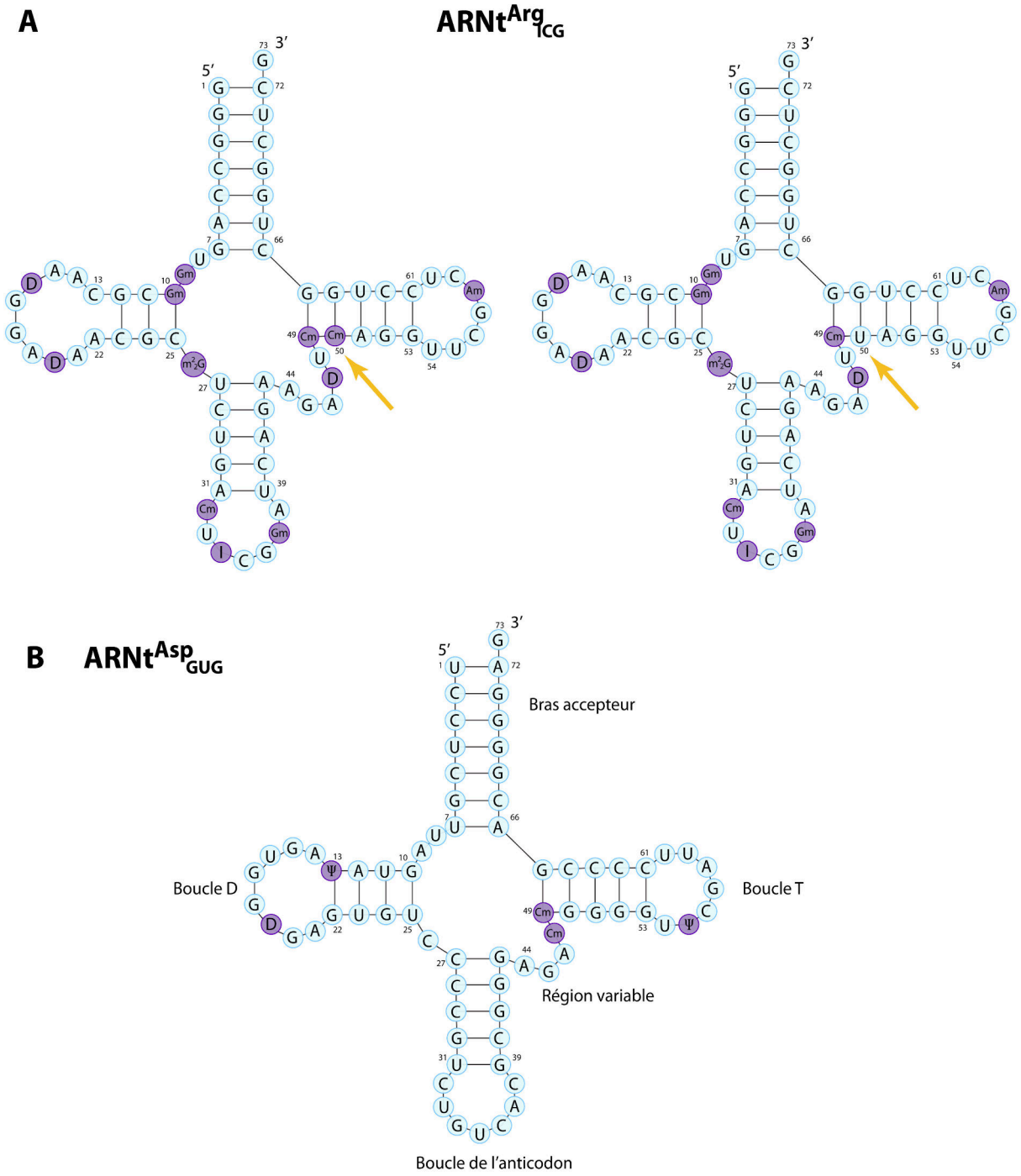



**Figure 30 : Résultats des analyses par spectrométrie de masse des (A) ARNt$^{Arg}_{ICG}$ et (B) ARNt$^{Asp}_{GTT}$.** Les modifications post-transcriptionnelles identifiées sont indiquées en violet. Xm : méthylation de la base X, D, dihydrouridine, Ψ, pseudouridine et I, inosine. La différence entre les deux ARNt$^{Arg}_{ICG}$ se situe au niveau de la modification qui est présente ou pas en position 50 et est indiquée par la flèche jaune.

Ainsi, seuls les $ARNt^{Arg}_{ICG}$ et $ARNt^{Asp}_{GTT}$ ont pu être utilisés pour déterminer leurs affinités relatives pour *Pf*tRip. Les faibles quantités d'ARNt purifiées n'ont pas permis d'utiliser la même approche que pour les transcrits. Dans le cas présent, les $ARNt^{Arg}_{ICG}$ et $ARNt^{Asp}_{GTT}$ purifiés ont été marqués radioactivement, et des concentrations croissantes (7 nM à 250 nM) d'ARNt total extrait des cellules HeLa ont été utilisées comme compétiteurs. Pour chaque expérience l'ARNt natif et le transcrit correspondant ont été testés en parallèle (Figure 31). La diminution de l'intensité des bandes ARNt lié ou tr lié indique donc le déplacement du complexe formé entre l'ARNt ou le transcrit avec *Pf*tRip par l'ARNt total. Plus le complexe disparaît vite, moins l'ARNt ou le transcrit sont de bons ligants pour *Pf*tRip.

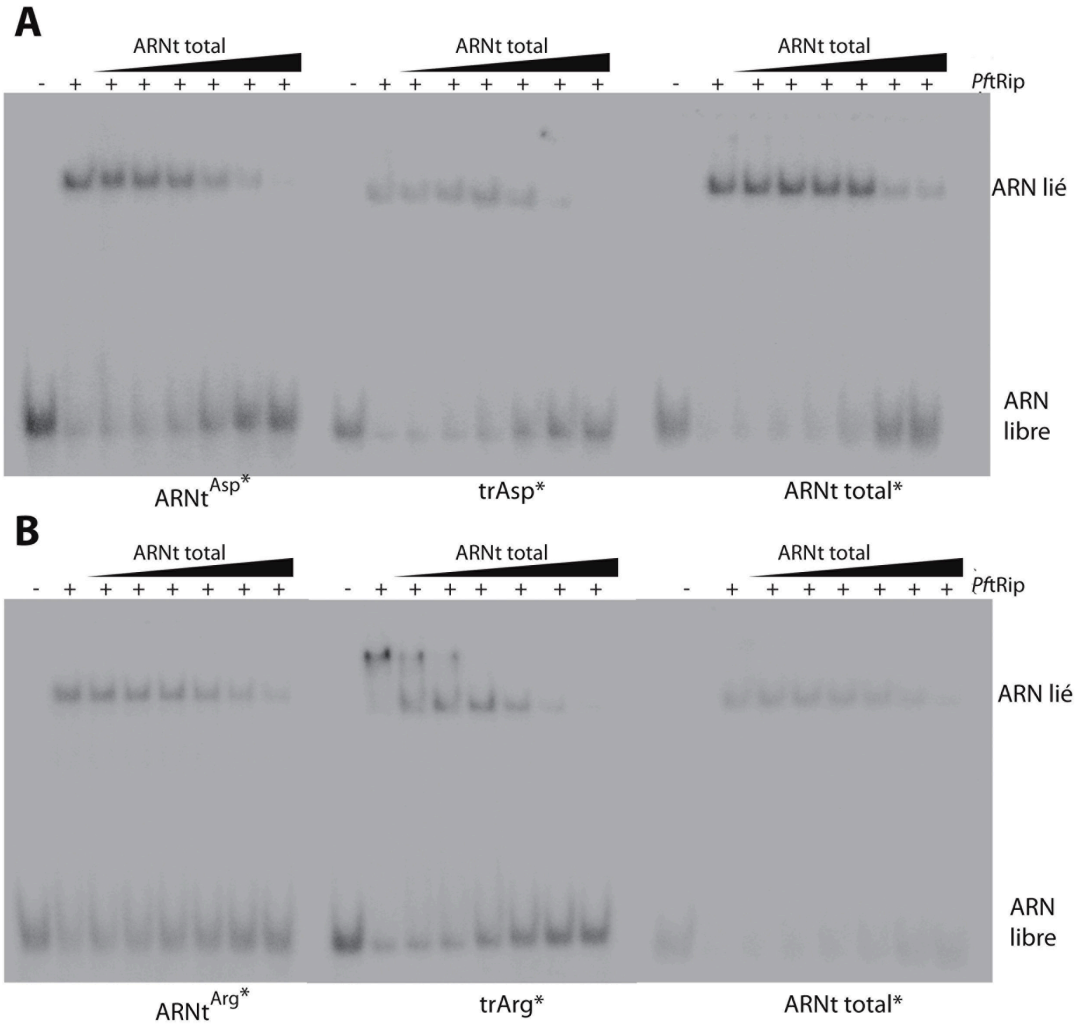



**Figure 31 : Expériences de retard sur gel avec (A) l'$ARNt^{Asp}$ et (B) l'$ARNt^{Arg}$ et les transcrits correspondants.** Sur chaque gel, le premier puits ne contient pas de protéine et sert de contrôle négatif, le deuxième puits contient uniquement l'ARNt ou le transcrit et *Pf*tRip, comme contrôle positif. Les 6 autres puits contiennent en plus des concentrations croissantes d'ARNt total compétiteur qui varient de 7 à 250 nM.

Les gels de compétition présentés dans la Figure 31A montrent que les ARNt$^{Asp}$ et tr$^{Asp}$ se comportent globalement de la même façon. En revanche dans la Figure 31B, on observe une meilleure résistance du complexe ARNt$^{Arg}$/*Pf*tRip en présence de l'ARNt total que du complexe tr$^{Arg}$/*Pf*tRip, suggérant que l'absence des modifications post transcriptionnelles affecte à la liaison de la protéine sur le transcrit. Cette différence de comportement est renforcée par la l'apparition du complexe O uniquement en présence de transcrit Arg (Figure 31B).

Les expériences ont été faites 4 fois pour chaque ARNt et chaque transcrit. Afin de comparer les affinités des ARNt et des transcrits pour *Pf*tRip toutes les bandes ont été quantifiées avec ImageJ et les données ont été traitées comme précédemment. Les Figures 32A et 32B mettent en évidence la proportion d'ARNt ou de transcrit lié à *Pf*tRip en fonction de la concentration en ARNt total. Alors que les transcrits Asp et Arg se comportent de la même façon (Figure 32A), les ARNt$^{Asp}$ et $^{Arg}$ montrent une différence significative dans leur capacité à lier *Pf*tRip (Figure 32B). En effet cette représentation confirme que le complexe *Pf*tRip/ARNt$^{Arg}$ se dissocie moins bien que le complexe *Pf*tRip/ARNt$^{Asp}$ en présence d'ARNt total.

Les efficacités de dissociation (E(diss)) calculées entre 31 et 125 nM d'ARNt total (Figure 32C) confirment aussi ces observations, montrant clairement que l'ARNt total déplace moins efficacement l'ARNt$^{Arg}$ que l'ARNt$^{Asp}$ ($p=0,0005$) et que le transcrit Arg ($p=0,01$) lié à *Pf*tRip, alors que les E(diss) de l'ARNt$^{Asp}$ et des deux transcrits Asp et Arg ne sont pas significativement différents ($0,15<p<0,77$) (Figure 32C).

L'ensemble de ces expériences indique que l'affinité de l'ARNt$^{Arg}_{ICG}$ pour *Pf*tRip est significativement supérieure à celle de l'ARNt$^{Asp}_{GTT}$, ce qui confirme les résultats obtenus précédemment par la technique de MIST (Tableau 3B). Dans les deux cas, on obtient une différence d'un facteur proche de 2 (1,7 pour le MIST et 2,2 pour le retard sur gel).

Même si notre incapacité à purifier d'autres ARNt ne nous permet pas de conclure définitivement, cette étude permet cependant de proposer que *Pf*tRip reconnaît les ARNt avec des affinités différentes, que l'approche MIST permet de les hiérarchiser et que cette discrimination ne dépend pas de la séquence de ces ARNt. En revanche, il semble que la présence de modifications post-transcriptionnelles joue un rôle essentiel dans cette discrimination.

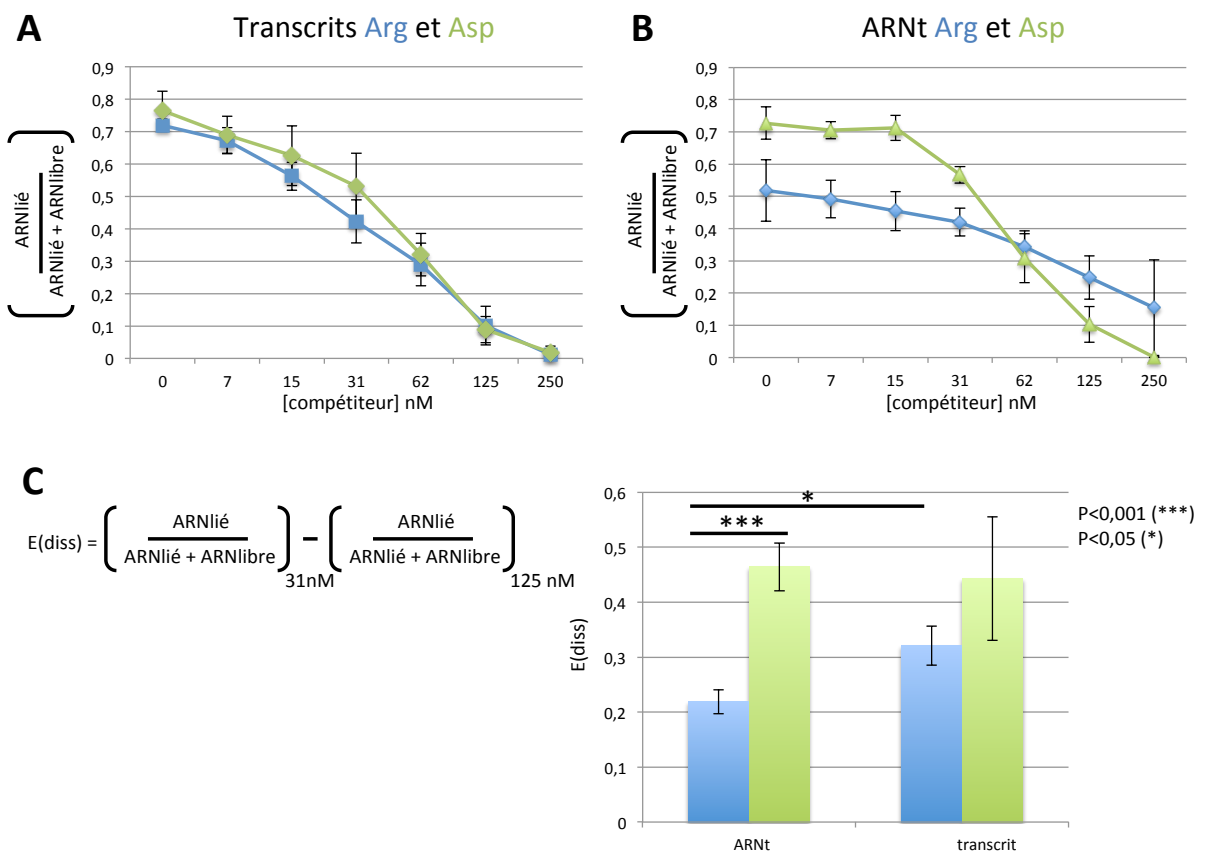



**Figure 32 : Analyse des résultats des compétitions entre les ARNt et les transcrits Asp (bleu) et Arg (vert) et l'ARNt total pour la liaison à *Pf*tRip. (A)** Dissociation des complexes transcrit/*Pf*tRip en présence d'ARNt total. **(B)** Dissociation des complexes ARNt/*Pf*tRip en présence d'ARNt total. **(C)** calcul et représentation des efficacités de dissociation (E(diss)). Les valeurs p ont été calculées par t-test, *** p<0.005 et * p<0,01. Chaque donnée est une moyenne de 4 expériences indépendantes.

### *1.7. Recherche d'un profil de modifications post transcriptionnelles modulant l'affinité de PftRip pour les ARNt*

Les modifications post-transcriptionnelles sont essentielles pour la stabilisation de la structure des ARNt (Revu dans Giegé, 2008, Motorin et Helm, 2010, Hopper, 2013). Elles sont, pour la plupart localisées au niveau du coude de la structure de l'ARNt et de la boucle anticodon. Dans leur article, Nomanbhoy, *et al.* (2001) ont déterminé les régions de l'ARNt reconnues lors de la formation du complexe avec un dimère de Trbp111 (Figures 33A) l'homologue bactérien de *Pf*tRip. Selon leur modèle, un des monomères de Trbp111 interagit avec les boucles D et T et la région variable, et l'autre monomère interagit avec la partie supérieure du bras anticodon de l'ARNt (Figure 33B). Nos propres expériences d'empreintes de *Pf*tRip sur l'ARNt$^{Asp}$) (Bour, *et al.*, 2016) confirment ces résultats.

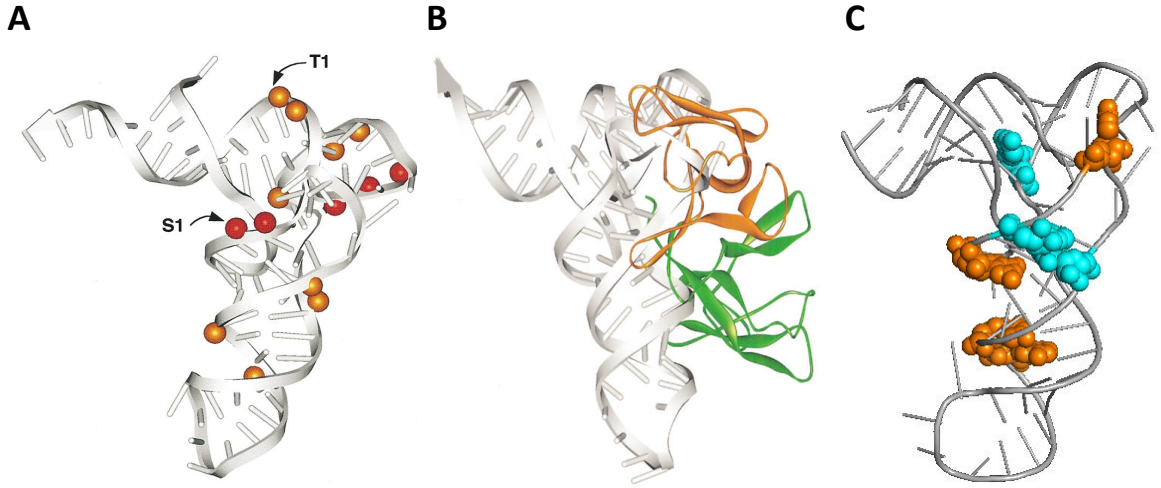



**Figure 33 : Reconnaissance des ARNt par Trbp111 (homologue de *Pf*tRip) et modifications post-transcriptionnelles. (A)** localisation des résidus de l'ARNt protégés en présence de Trbp111. **(B)** Modèle de l'intraction entre l'ARNt et un dimère de Trbp111. **(A)** et **(B)** sont adaptés de (Nomanbhoy, T., et al., (2001). **(C)** récapitulation des positions modifiées pouvant intervenir dans la discriminatif positive (orange) ou négative (bleue) des ARNt par *Pf*tRip.

Nous avons donc recherché dans la banque de donnée Modomics (http://modomics.genesilico.pl/) tous les ARNt cytosoliques de vertébrés dont les modifications post-transcriptionnelles ont été identifiées et qui ont été testés dans le MIST. Nous avons récupéré 21 séquences qui représentent bien les différentes affinités mesurées par MIST et répertorié les modifications localisées dans le bras D, le bras anticodon, la région variable et la boucle T (Tableau 6). Parmi ces modifications, il ressort que certaines pourraient favoriser ou réduire l'efficacité de reconnaissance entre les ARNt et *Pf*tRip. Par exemple, une pseudouridine (Ψ) en position 13 dans le bras D, une dihydrouridine (D) ou une Ψ en position 20b dans la boucle D, ou encore un 5-methylcytidine ($m^5C$) ou une Ψ en position 50 dans la région variable pourraient inhiber l'interaction avec *Pf*tRip. Au contraire, un N4-acetylcytidine ($ac^4C$) en position 12 dans le bras D, une methylguanosine (Gm) (Tarun, *et al.*, 2008) en position 18 dans la boucle D ou une Ψ en position 27 et/ou 28 dans le bras anticodon pourraient favoriser cette interaction.

Lorsque ces positions sont localisées sur la structure 3D d'un ARNt (Figure 33 C), on observe que tous les résidus concernés sont superposables aux protections laissées par Trbp111 sur l'ARNt (comparer Figure 33 A et 33C).

**Tableau 6 : Récapitulatif des modifications post-transcriptionnelles et affinité des ARNt pour *Pf*tRip.**

| | | Bras D | | | Boucle D | | | | | | | | Bras Ac | | | Région variable | | | | | Boucle T | | | Espèce |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Sonde** | **ARNt** | **9** | **10** | **12** | **13** | **14** | **16** | **17** | **18** | **20** | **20a** | **20b** | **26** | **27** | **28** | **46** | **47** | **48** | **49** | **50** | **54** | **55** | **58** | |
| **Asp-1h** | AspGTC | | | | Ψ | | | | | D | | | | | | | | m5C | m5C | | T | Ψ | | cette étude (*R.norvegicus*) |
| **Arg-1h** | ArgiCG | m1G | m2G | | | | D | | | | D | | m2,2G | | | | D | | m5C | | | Ψ | m1A | cette étude |
| **Glu-1h** | GluyTC | | m2G | | Ψ | | | | | D | Ψ | | | | | | | | m5C | m5C | m5Um | Ψ | | *H.sapiens CTC* |
| **Tyr-2h** | TyrGTA | | m2G | | | | D | D | | acp3U | | | m2,2G | m2,2G | | m7G | D | m5C | | | T | Ψ | m1A | *H.sapiens GUA* |
| **Leu-3h** | LeuCAA | | m2G | ac4C | | | | | | D | | Ψ | m2,2G | | | | | m5C | | | T | Ψ | m1A | *H.sapiens GAA* |
| **Glu-2h** | GluTTC-1 | m1G | | | Ψ | | | | | D | D | | | | | | | | m5C | m5C | m5Um | Ψ | | *R.norvegicuus UUC* |
| **Gly-1h** | GlysCC | | | | | | | | | D | | D | | | | | | m5C | m5C | m5C | T | Ψ | m1A | *H. sapiens GCC* |
| **Val-1h** | ValmAC | | | | Ψ | | D | D | | D | D | D | m2G | Ψ | | m7G | D | m5C | m5C | | | Ψ | m1A | *H.sapiens CAC/M.musculus AAC* |
| **Arg-2h** | ArgyCG | m1G | m2G | | | | D | | Gm | | D | | | Ψ | | | D | | | Ψ | T | Ψ | m1A | *B.taurus CCG* |
| **Met-2h** | Met-e | | m2G | | | | D | | | D | | | m2G | Ψ | | m7G | D | m5C | | | T | Ψ | m1A | *H.sapiens CUA* |
| **His-1h** | HisGTG | | | | Ψ | | D | | | D | D | | | | | | | m5C | m5C | | | Ψ | m1A | *H.sapiens GUG* |
| **Thr-3h** | ThrCGT-1 | m1G | m2G | | | | D | | | D | | | m2,2G | | | | D | m5C | | | | Ψ | m1A | *B.taurus* |
| **Met-1h/Thr6h** | Met-i | m1G | m2G | | | | | | | | | | m2G | | | m7G | D | m5C | | | | | m1A | *H.sapiens ini* |
| **Arg-3h** | ArgCCT | m1G | m2G | | | | D | | | | D | | | | | | D | | | | T | Ψ | m1A | *B.taurus CCU* |
| **Phe-1h** | PheGAA | | m2G | | | m1A | D | D | | | | | m2,2G | Ψ | Ψ | m7G | D | | m5C | | T | Ψ | m1A | *H.sapiens GAA* |
| **Ser-3h** | SerGCT | | | ac4C | | | D | | Gm | D | D | | m2,2G | | Ψ | | | m5C | | | T | Ψ | m1A | *R.norvegicus GCU* |
| **Arg-4h** | ArgTCT | m1G | | | | | | | | | m1G | | | Ψ | Ψ | m7G | D | | m5C | | | | | *B.taurus UCU* |
| **Ser-2h** | SerwGA | | | ac4C | | | D | | Gm | D | D | | | | Ψ | | | m5C | | | T | Ψ | m1A | *H.sapiens UGA* |
| **Leu-2hn** | LeuCAG | | m2G | ac4C | | | | | | D | D | | m2,2G | | | | | m5C | | | T | Ψ | m1A | *B.taurus CAG* |
| **Asn-1h** | AsnGTT | m1G | m2G | | | | D | | | D | acp3U | | m2,2G | Ψ | Ψ | m7G | D | | | | ? | Ψ | m1A | *H.sapiens GUU* |
| **Ser-4h** | SerAGA | | | ac4C | | | D | | Gm | D | D | | m2,2G | | Ψ | | | m5C | | | T | Ψ | m1A | *R.norvegicus AGA* |

Dans l'état actuel de nos expériences, nous ne pouvons pas conclure définitivement quant au rôle des modifications post-transcriptionnelles sur la reconnaissance des ARNt par *Pf*tRip. Par exemple, il est difficile de savoir si ces modifications ont un effet direct ou indirect sur la liaison de *Pf*tRip. Afin de confirmer ces résultats, nous envisageons d'utiliser des ARNt de levure purs contenant des jeux de modifications différents pour tester leur capacité à se lier à *Pf*tRip et voir si les résultats confirment nos prédictions établies par MIST. Dans ce cas, nous utiliserons deux approches : les retards sur gel et la titration calorimétrique isotherme (ITC). Cette dernière consomme beaucoup de matériel mais permet aussi de déterminer des constantes de dissociation plus précises que par la technique du retard sur gel.

### *1.8. Conséquences sur la biologie du parasite*

Enfin, en nous basant sur le classement des affinités des ARNt pour *Pf*tRip déterminé par nos expériences de MIST, nous nous sommes demandé si il y avait une corrélation entre cette affinité, donc la probabilité de chaque ARNt à être importé, avec l'usage de ces ARNt dans la traduction des protéines du parasite. Pour cela nous avons déterminé les codons reconnus par chaque ARNt et calculé la proportion de codons potentiellement décodés par chaque ARNt testé. Le diagramme de la Figure 34 montre qu'il n'y a aucune corrélation globale entre l'affinité de *Pf*tRip pour un ARNt donné et son utilisation putative dans la traduction des protéines de *Plasmodium*.

Cependant, Il est intéressant de noter que l'ARNt$^{Asn}$ fait partie des meilleurs ligands de *Pf*tRip. Or, l'asparagine (asn) est aussi l'acide aminé le plus utilisé dans le protéome du parasite. Non seulement l'asn est l'acide aminé le plus abondant (avec la lysine), mais surtout il n'est pas réparti de façon homogène dans les protéines. Au contraire, on retrouve la majorité des asn dans des insertions de faible complexité et ces asn peuvent former de longues répétitions. Les insertions riches en asn sont présentes dans toutes les protéines de *Plasmodium* (*falciparum, berghei*…), à l'exception des protéines ribosomiques. Leur rôle est encore très discuté. Dans sa revue, Davies, *et al.*, (2017) font un récapitulatif des nombreux processus dans lesquelles ces répetitions sont soupçonnées d'être impliquées, comme la stabilisation de l'ARNm (Xue et Forsdyke, 2003), la fonctionnalité (Ferguson, *et al*., 2014), (Chaudhry, *et al*., 2018) et la localisation (Davies, *et al.*, 2016) des protéines, la virulence du parasite (Cutts, *et al.*, 2017) ou encore l'immunité (Kemp, *et al.*, 1987). Notre équipe a

proposé une hypothèse où l'accumulation des codons Asn dans les insertions des protéines de *Plasmodium* permettrait de moduler localement la vitesse de traduction de ces protéines en ralentissant le ribosome lorsque la concentration en Asn-ARNt$^{Asn}$ est limitante dans le parasite (Frugier, *et al.*, 2010, Filisetti, *et al.*, 2013).

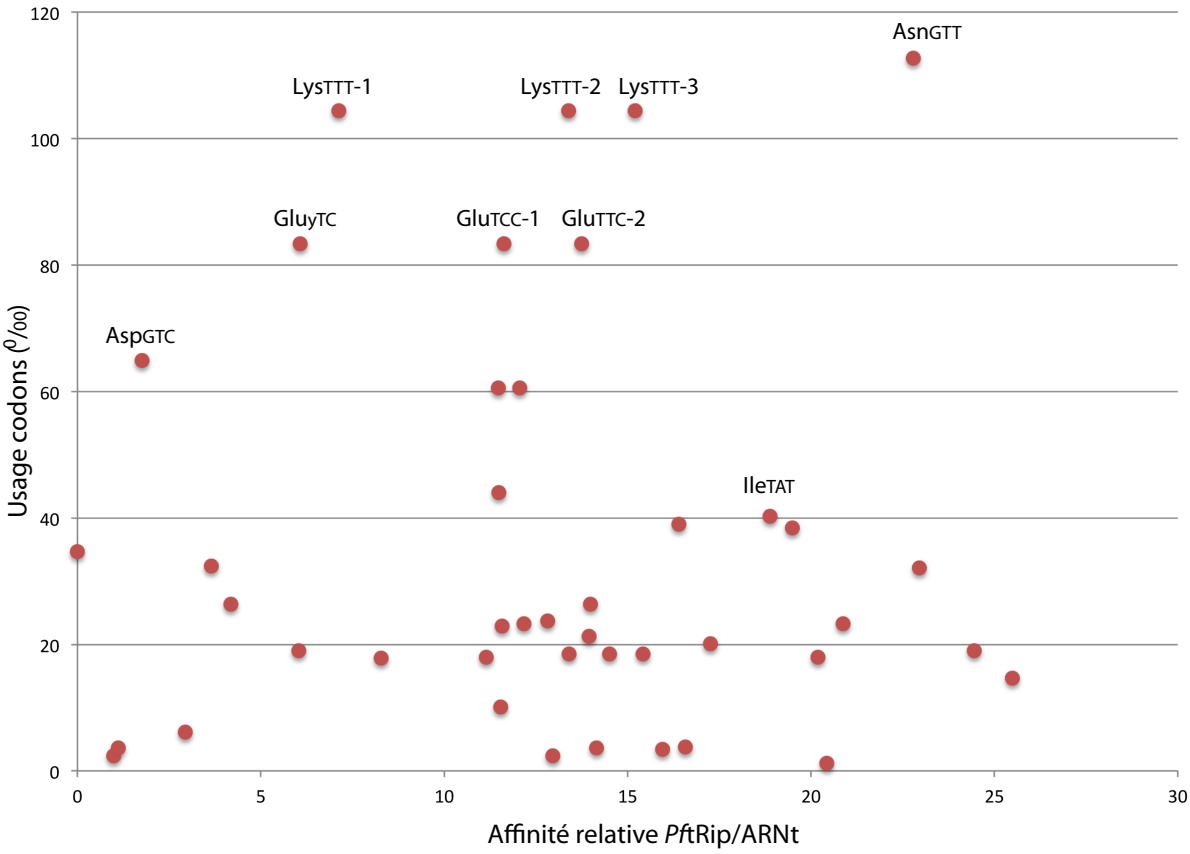



**Figure 34: Usage potentiel des ARNt dans le parasite en fonction de leur affinité pour *Pf*tRip.** Chaque point correspond à un ARNt et est défini par son affinité pour *Pf*tRip définie par MIST (abscisse) et par son usage dans la traduction des protéines de *Plasmodium* (ordonnée). Les ARNt correspondants aux acides aminés les plus utilisés par le parasite sont indiqués.

De plus, des expériences de protéomiques comparatives entre le parasite sauvage et le parasite tRip-KO ont été faites dans le laboratoire. Elles montrent qu'une grande partie des protéines dont l'expression est réduite dans le parasite tRip-KO sont caractérisées par leur richesse en asparagine, suggérant que le parasite KO ne produirait pas suffisamment d'Asn-ARNt$^{Asn}$ pour couvrir tous ses besoins. Ainsi, en important des ARNt$^{Asn}$ de l'hôte, le parasite pourrait compenser une concentration endogène trop faible à certaines étapes de son cycle de développement.

Cependant, le protéome du parasite tRip-KO est caractérisé par la sous-expression d'autres protéines impliquées dans l'invasion ou la virulence alors que celles-ci ne contiennent pas plus de 10 % d'asn. Cette observation suggère qu'en dehors de la traduction, les ARNt importés pourraient également être impliqués dans des processus encore inconnus de signalisation et de régulation de l'expression des protéines. De tels mécanismes ont été mis à jour dans d'autres organismes (Revu dans Schimmel, 2017 et Kirchner, 2015). Les ARNt ont été associés à des voies de signalisation impliquées dans des processus très différents, dont la résistance des bactéries à certains antibiotiques (Roy et Ibba, 2008), à la transcription inverse chez les virus (Ruggero, *et al*., 2014), à la synthèse de l'hème et de la chlorophylle (Jahn, *et al.*, 1992). Chez l'Homme, l'abondance d'ARNt$^{Glu}{}_{UUC}$ et $^{Arg}{}_{CCG}$ a été associée à la voie de signalisation responsable de métastases dans les tumeurs du sein (Goodarzi, *et al.*, 2016). L'implication des ARNt dans la réponse au stress a aussi été décrite (Hopper, 2013), notamment, la voie mTOR, qui se met en place lors d'un choc thermique pour délocaliser les ARNt et ainsi activer leur dégradation (Watanabe, *et al*., 2014). En plus, les rôles des fragments d'ARNt sont de plus en plus étudiés. Ils sont impliqués dans la régulation épigénétique (Chen, *et al.*, 2016) et dans la réponse au stress (Durdevic, *et al*., 2013).

L'ensemble de ces observations supporte notre hypothèse selon laquelle les ARNt importés par le parasite ne seraient pas uniquement utilisés pour la traduction. En effet, certains ARNt, dans le cas où ils ne seraient pas reconnus par les aaRSs de *Plasmodium* pourraient participer à la régulation de l'expression génique du parasite et à son adaptation aux différents environnements cellulaires rencontrés au cours de son développement.

## 2. Le SELEX

L'étude de la spécificité de *Pf*tRip a été également abordée en sélectionnant des molécules d'ARN qui se lient à la protéine dans des conditions expérimentales de plus en plus stringentes. Pour ce faire, la technique SELEX (Systematic Evolution of Ligands by Exponencial enrichment) a été utilisée. Cette technique est basée sur la capacité d'un aptamère à se lier à la protéine cible. C'est une technique puissante qui permet d'identifier des motifs nucléotidiques ou des motifs structuraux, spécifiques pour la reconnaissance par une protéine donnée.

### *2.1. Préparation du matériel*

La bibliothèque d'aptamères initiale, contenant 25 nucléotides aléatoires (Figure 35A), a été construite par PCR et transcrite *in vitro*, comme indiqué dans la partie « matériel et méthodes ». Sept nmoles d'ARN (71 nucléotides) ont été obtenus après transcription *in vitro* et purification sur gel (Figure 35B). Cela correspond à 4,2 x $10^{15}$ molécules et coïncide avec la complexité théorique de la bibliothèque initiale (les 25 nucléotides couvrent 1,12 x $10^{15}$ possibilités). Chaque séquence est donc théoriquement représentée au moins trois fois.

**A**

**MATRICE**

```
5'-TAATACGACTCACTATAGGGAGACAAGAATAAACGCTCAA-3'                    3'-TTCGACAGGAGGCTCACAACAGGC-5'
                    5'-GGGAGACAAGAATAAACGCTCAANNNNNNNNNNNNNNNNNNNNNNNNNNNTTCGACAGGAGGCTCACAACAGGC-3'
```

**PRODUIT PCR**

```
5'-TAATACGACTCACTATAGGGAGACAAGAATAAACGCTCAANNNNNNNNNNNNNNNNNNNNNNNNNNNTTCGACAGGAGGCTCACAACAGGC-3'
3'-ATTATGCTGAGTGATATCCCTCTGTTCTTATTTGCGAGTTNNNNNNNNNNNNNNNNNNNNNNNNNNNAAGCTGTCCTCCGAGTGTTGTCCT-5'
```

**ARN**

```
5'-CCCTCTGTTCTTATTTCGTATTNNNNNNNNNNNNNNNNNNNNNNNNNNNAAGCTGTCCTCCGAGTGTTGTCCG-34
```

**B**

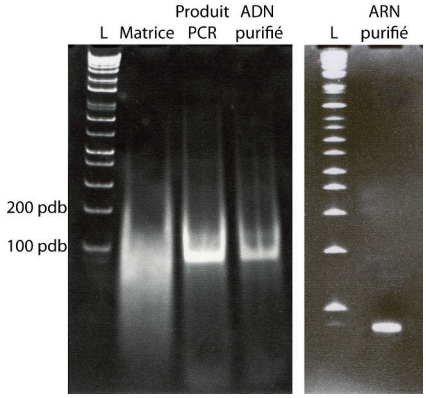



**Figure 35 : Production de la bibliothèque d'ARN. (A)** La matrice simple brin a été amplifiée par PCR grâce aux deux amorces, l'ADN double brin produit a été purifié sur gel natif avant d'être transcrit en ARN *in vitro* par l'ARN polymérase T7. **(B)** Gels analytiques : Les types de molécules produites au cours des différentes étapes de la production de la bibliothèque initiale d'ARN ont été analysées sur gel natif (à gauche) et dénaturant (à droite). L correspond au marqueur de poids moléculaire.

Dans cette étude nous avons fait le choix d'utiliser uniquement le domaine C-terminal (*Pf*tRip$_{214-402}$), seul responsable de la liaison aux ARNt dans *Pf*tRip (Figure 36A). Trois protéines ont été utilisées au cours des différentes étapes de sélection (Figure 36B) : la protéine GST (contrôle négatif), le domaine C-terminal de *Pf*tRip fusionné à la GST (GST-

*Pf*tRip$_{214-402}$), et la fusion GST-*Pf*tRip$_{214-402}$Mut où le site de reconnaissance de *Pf*tRip pour l'ARNt a été muté (Ser$_{312}$ en Ala et Met$_{312}$ en Ala). Ce mutant ne se lie plus aux ARNt.

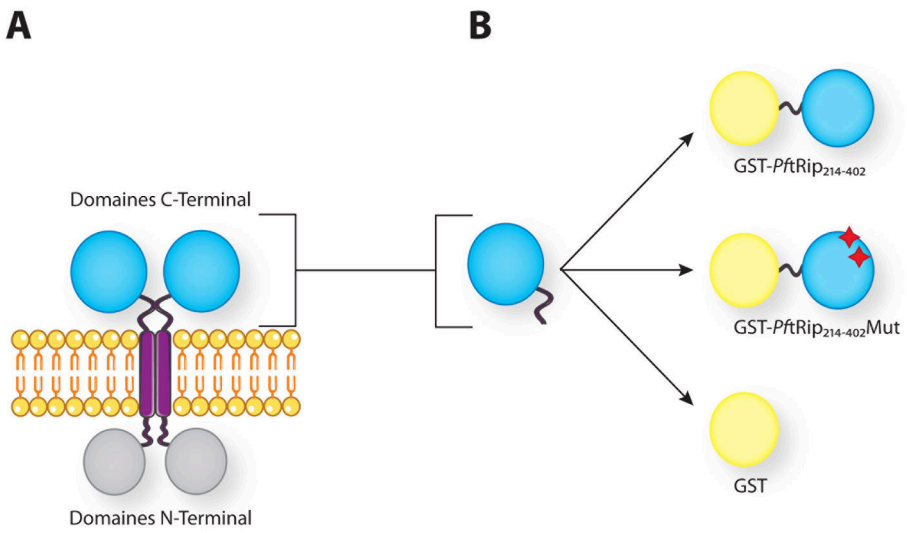



**Figure 36 : Schéma représentant *Pf*tRip et les différentes protéines utilisées lors des étapes de sélection. (A)** Les domaines N-terminal et C-terminal de *Pf*tRip sont représentés par des sphères grise et bleue, respectivement alors que le domaine transmembranaire est indiqué en violet. Seul le domaine C-terminal est impliqué dans la reconnaissance des ARNt. **(B)** Les protéines de fusion ont été construites en fusionnant le domaine GST à l'extrémité N-terminal du domaine de liaison aux ARNt de *Pf*tRip. Dans ces constructions, la GST remplace le domaine N terminal et le domaine transmabranaire de *Pf*tRip. Les étoiles rouges indiquent les deux mutations présentes dans le site de liaison aux l'ARNt (résidus sérine$_{312}$ et méthionine$_{315}$ mutés en alanine) dans le domaine C-terminal de *Pf*tRip.

GST-*Pf*tRip$_{214-402}$ a été utilisée pour les sélections positives, afin de retenir tous les aptamères qui se lient au domaine C-terminal de *Pf*tRip. Les deux autres protéines (GST et GST-*Pf*tRip$_{214-402}$Mut) ont été utilisées pour augmenter la stringence des sélections en utilisant des sélections négatives et éliminer les aptamères qui se lient soit à la partie GST de la protéine de fusion soit à une région autre que le site de reconnaissance des ARNt du domaine C-terminal de *Pf*tRip.

Les trois protéines ont été purifiées sur une colonne de glutathion-agarose. La Figure 37 montre la pureté des protéines utilisées dans les tests de sélections. Une purification à partir de 250 ml de culture a généré 5,5 µmol des deux versions du domaine C-terminal de *Pf*tRip (GST-*Pf*tRip$_{214-402}$ et GST-*Pf*tRip$_{214-402}$Mut) et 400 µmol de GST, quantités suffisantes pour mener à bien cette étude.

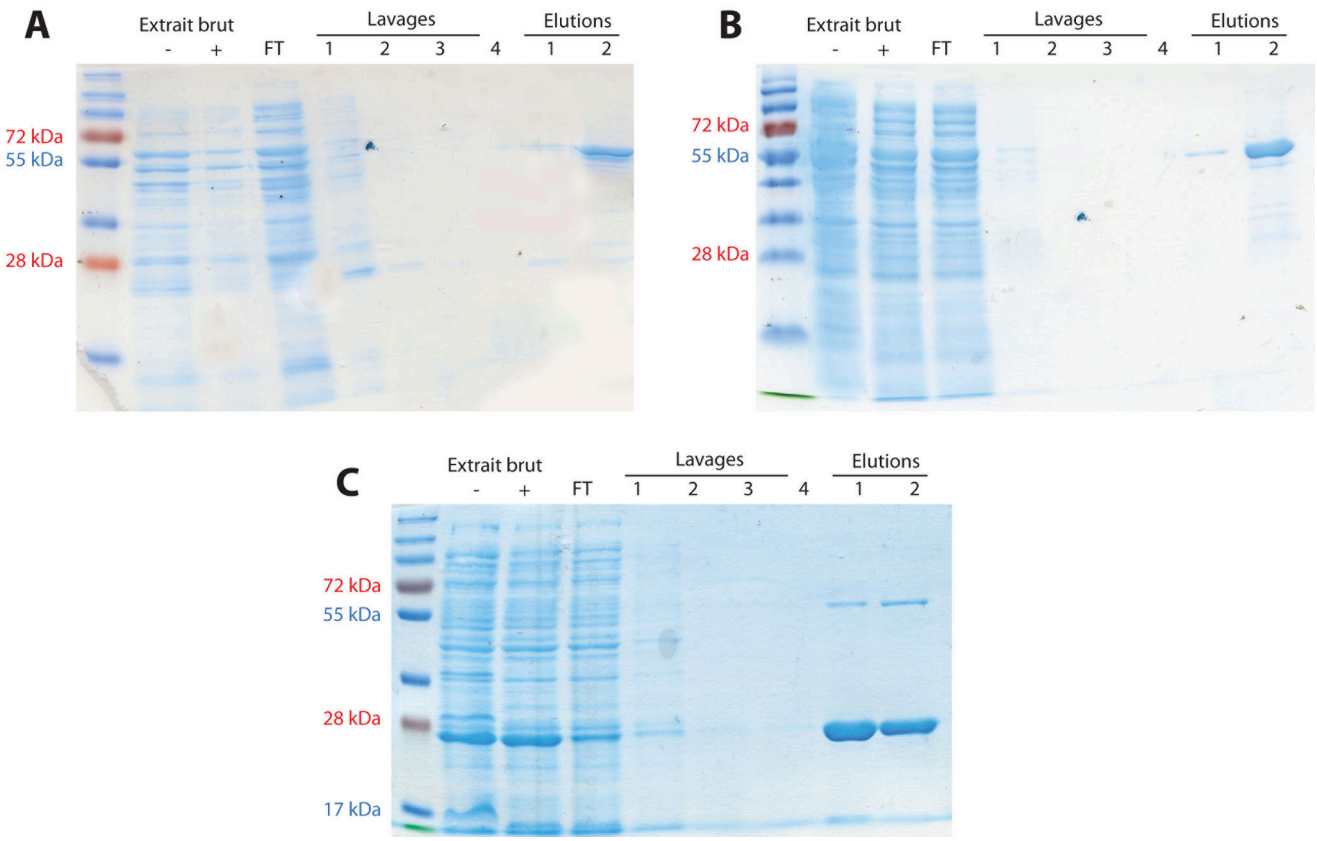



**Figure 37: Production des protéines (A) GST-*Pf*tRip$_{214\text{-}402}$, (B) GST-*Pf*tRip$_{214\text{-}402}$Mut et (C) GST.** Les protéines ont été purifiées sur colonne glutathion-agarose en batch ; deux échantillons correspondent à l'extrait brut avant (-) et après (+) centrifugation. FT correspond à la fraction des protéines qui ne restent pas fixées sur la résine. Les protéines sont éluées en présence de 10 mM glutathion et migrent sur gel aux tailles attendues (25 kDa pour la GST et 55 kDa pour les deux protéines de fusion).

### *2.2. Les cycles de sélection*

Les 5 étapes de sélection sont récapitulées dans le Tableau 7. Chaque sélection a été réalisée en faisant varier les conditions de temps (entre 5 et 30 min d'incubation), de température (entre 26 et 37°C) et de concentration en ARN (de 7 à 0,45 µM). A partir de la deuxième étape, une sélection négative a systématiquement été effectuée en amont de la sélection positive pour éliminer le plus grand nombre d'aptamères non spécifiques.

**Tableau 7 : Récapitulatif des conditions de sélection.**

| Etape | ARN initial | Sélection négative | Sélection positive |
|---|---|---|---|
| 1 | 7000 pmol | | 26°C<br>10 min<br>320 pmol |
| 2 | 5000 pmol | 26°C<br>10 min<br>320 pmol | 26°C<br>30 min<br>320 pmol |
| 3 | 2000 pmol | 26°C<br>30 min<br>320 pmol | 26°C<br>30 min<br>320 pmol |
| 4 | 450 pmol | 37°C<br>5 min<br>320 pmol | 37°C<br>5 min<br>320 pmol |
| 5 | 450 pmol | 26°C<br>30 min<br>320 pmol | 37°C<br>5 min<br>320 pmol |

Le tableau récapitule les quantités initiales d'ARN et de protéines, les temps et les températures d'incubation au cours des différentes sélections. Se reporter à la légende de la Figure 36 pour les couleurs.

Pour l'ensemble des sélections, chaque étape a été contrôlée sur gel analytique dénaturant : sélection négative (FT, élution), sélection positive (lavages et élution finale). Sur le gel de la 38, correspondant à la deuxième étape de sélection, on observe que la sélection négative permet d'éliminer une quantité significative d'ARN qui se lient de façon aspécifique sur la résine (élution/sélection négative). Seul l'ARN présent dans la fraction qui ne se fixe pas sur les différents supports de sélections négatives (résine, GST ou GST-*Pf*tRip$_{214-402}$Mut = FT, Figure 38) est utilisé dans la sélection positive correspondante en présence de GST-*Pf*tRip$_{214-402}$ immobilisée (0,32 µM). La résine est lavée jusqu'à ce que l'absorbance redescende à zéro. La dernière étape consiste à éluer les complexes ARN/GST-*Pf*tRip$_{214-402}$ en

présence de glutathion (10 mM). La présence d'ARN sélectionnés est confirmée par une bande plus ou moins intense sur le gel (Elution).

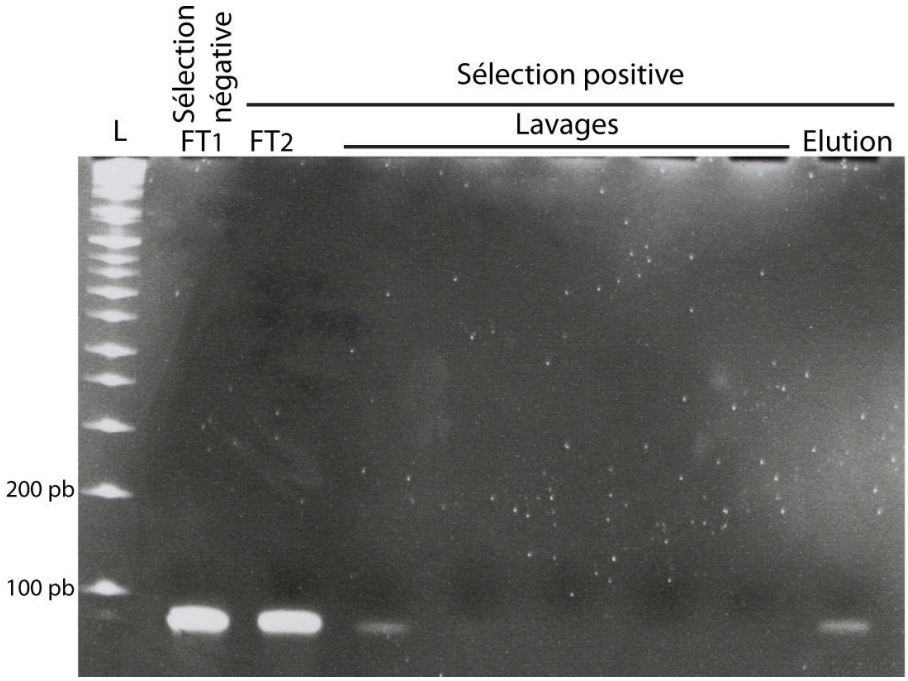



**Figure 38 : Gel analytique récapitulant les différentes étapes d'une sélection (Sélection #2).** Seuls les aptamères ne se fixant pas sur la résine (fraction FT) au cours de la sélection négative sont utilisés dans la sélection positive. La résine est lavée abondamment et les ARN fixés sur la protéine GST-*Pf*tRip$_{214-402}$ sont élués en présence de glutathion. Les ARN sont séparés de GST-*Pf*tRip$_{214-402}$ par un traitement au phénol avant d'être réverse transcrits, amplifiés par PCR et transcrits *in vitro* pour la sélection suivante.

Le Tableau 8 montre l'efficacité d'enrichissement des 5 étapes de sélection. Logiquement on observe une faible efficacité de sélection au cours des deux premières étapes, indiquant qu'une faible proportion des aptamères est capable de se lier au domaine C-terminal de *Pf*tRip. Dans un deuxième temps, cette efficacité de sélection augmente, jusqu'à atteindre un plateau de 13 %, suggérant que les ligands spécifiques les plus efficaces ont été sélectionnés.

Les ARN obtenues après la cinquième étape de sélection ont été amplifiés et clonés dans pDrive. Parmi les 50 clones analysés, 28 contenaient un insert et ont été séquencés (Figure 39). Chaque séquence est unique, et leur alignement ne nous a pas permis de mettre en évidence de motifs nucléotidiques communs.

**Tableau 8 : Efficacité des étapes de sélection**

| Etape | ARN (µM) | [ARN]/ [GST-PftRip214-402] | ARN sélectionnés (pmoles) | Efficacité de sélection (%) |
|---|---|---|---|---|
| 1 | 7 | 22-janv | 73 | 1,04 |
| 2 | 5 | 15-janv | 43,5 | 0,87 |
| 3 | 2 | 06-janv | 49 | 2,44 |
| 4 | 0,45 | 1,5/1 | 62 | 13,8 |
| 5 | 0,45 | 1,5/1 | 57,5 | 12,78 |

Les sélections ont été effectuées dans un volume constant de 1 ml en présence de 320 pmoles de GST-*Pf*tRip$_{214-402}$ (0,32 µM).

```
ARN30    --GGTACAGCCGTCCATC--AGAAACCTA----
ARN47    --ACGGTAGCCAACCCAT--CGAAACCTA----
ARN17    ---GGTGAGATAACCTGT--CAGCACCTAG---
ARN13    -----ATAGTGGACCTGTTTCGAAACCTA----
ARN28    ---GACGAGTCCAAAAGTT-CGATGCCTA----
ARN44    -------CTTCCTTCAAAT-CGCTGCTAGCAGT
ARN32    -----CTAGAGACACCCCGACTCCGCCCCA---
ARN48    -----CTCGAGACGCATCTCCACCGCCCCG---
ARN14    -----CTAGAGACATGCTGACCCCTCGGAC---
ARN42    -----TTAGAGACACTCTCGGGGCATCCA----
ARN46    -----CCAGTGACCTCTTCGCCACACCCAG---
ARN24    ------GAGAAGTCAGATCGAAACACCTAGA--
ARN15    -----ATGAACACCTAGTCAAGACCGTTTA---
ARN33    ----CACGT-CACCTAGACATGAACGTATA---
ARN27    ---TACTGCACACCTAGTGTAAATCTTA-----
ARN16    -----AGCTTTACCTAAAGGCATTAGTAGA---
ARN37    ------ACCTGACCTACAGGTAACGCTAACA--
ARN23    ------TTCTGTCCTCATCATAGCGTTACTA--
ARN29    -------GCTGCGCTTAACAGCTAGTCACTTA-
ARN45    ------AAACCAACTGACCCAGACGGCGACTA-
ARN18    TCTTCCGGGTTCCCCTCAAATATCGG-------
ARN19    ----CCTTGTCCCGCTCCCATATCCTTGT----
ARN26    ----CCGAAGCCCCCTTCACTACCTAAA-----
ARN35    ----TCTCTTTCCCCTTACTCCCCATTCT----
ARN43    ----TCACCTTGCCATTCCCCGCAGTTTA----
ARN31    -----CCAATCCCTTTGGCAAGCAGCGGTA---
ARN40    ------AACGTGTGTAGGTAAACAATGTGAG--
ARN34    -----GGAGTCATTTACTGCATCGGACAAT---
```

**Figure 39 : Alignement des séquences des aptamères obtenus après 5 étapes de sélection** (CLUSTAL 2.1 multiple sequence alignment). Les résidus pyrimidiques sont indiqués en rouge. Seuls les nucléotides correspondant à la séquence sélectionnée de l'aptamère sont représentés.

### *2.3. A la recherche d'un motif commun*

Les séquences de ces aptamères sont remarquables quant à la forte proportion de pyrimidines qu'elles contiennent (Figure 39). L'utilisation du logiciel RNAfold a permis de prédire des structures secondaires (Figure 40) qui ont toutes en commun que la séquence sélectionnée interagit avec les séquences imposées lors de la conception de la librairie.

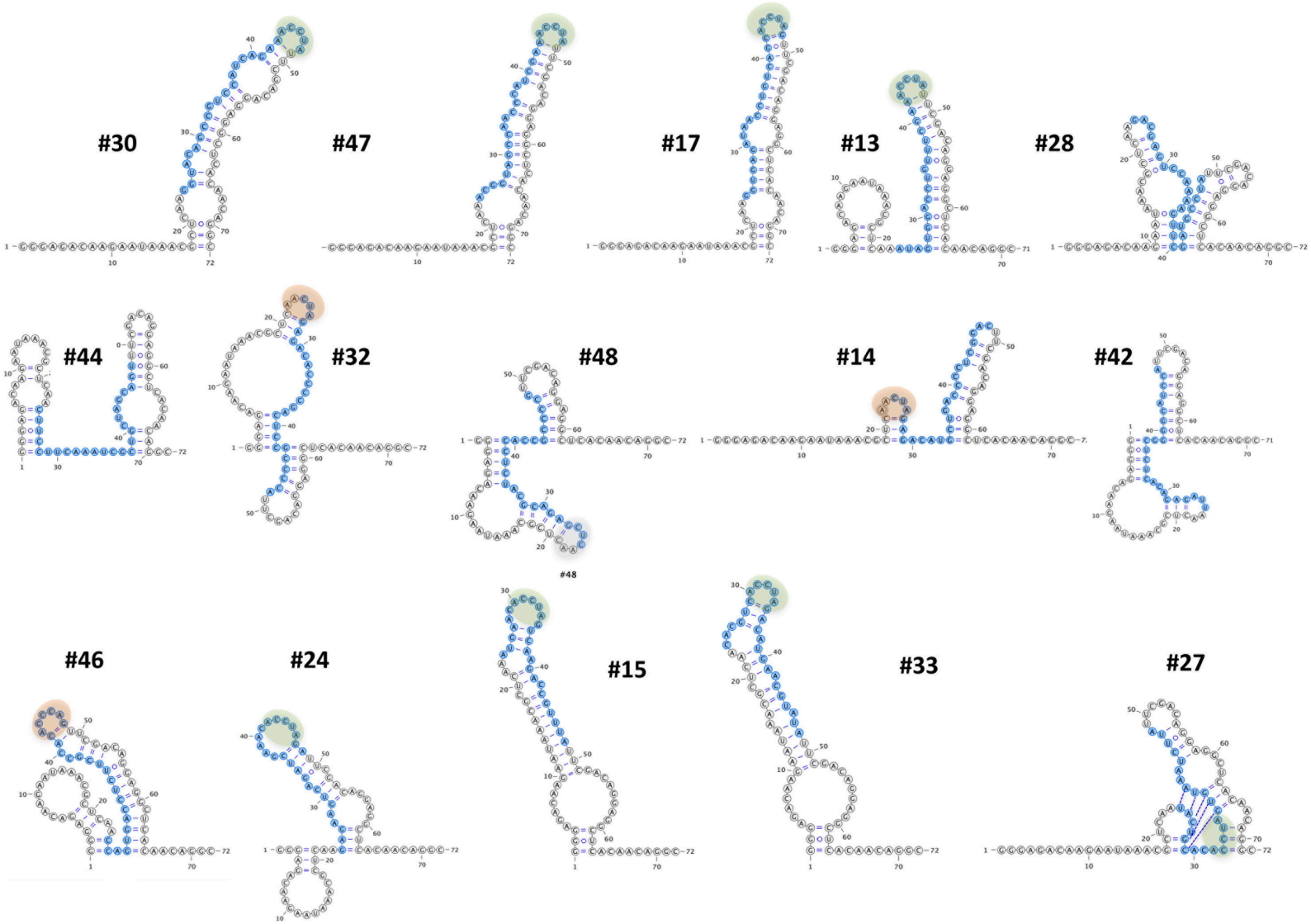
Grace à ces prédictions de structure secondaire, nous avons pu identifier un motif conservé dans la moitié des ARN sélectionnés. Il s'agit de la séquence 5'-ACCUA-3' présente dans une boucle apicale. Cette séquence est strictement conservée dans les aptamères 30, 47, 17, 13, 24, 15, 33, 16, 37, 26 et 27. Les ARN 32, 14, 45, 43 et 46 présentent également cette séquence dans une boucle apicale avec cependant une mutation ponctuelle en position 2, 4 ou 5. Pour ce qui est des autres aptamères, aucun consensus évident n'est détectable.

### *2.4. Comparaison des affinités des aptamères*

Les affinités de chacun des aptamères pour *Pf*tRip ont été comparées par des expériences de compétition. Dans ces expériences, un complexe initial est formé entre *Pf*tRip (entière) et l'ARNt$^{Asp}$ de levure marqué radioactivement. Deux concentrations d'aptamères (40 et 160 nM) ont été testées afin de déterminer leur capacité à dissocier le complexe préétabli (Figure 41). Tous les aptamères ne dissocient pas le complexe *Pf*tRip/ARNt avec la même efficacité. Pour certains, on ne distingue pas de diminution de la radioactivité au niveau de l'ARNt lié et/ou d'augmentation dans la bande correspondant à l'ARNt libre, deux situations qui indiquent que le compétiteur déplace l'ARNt$^{Asp}$ pour se lier à sa place sur *Pf*tRip. Il se peut que l'absence d'effet observée pour certains aptamères soit la conséquence de concentrations trop faibles dans le test pour entraîner une dissociation.

En revanche, sur les 28 séquences testées, les aptamères 26, 15, 32, 13, 16, 17, 24, 27 tous porteur de la séquence ACCUA et les aptamères 23, 35, 43, 44, 42, 28 et 29 (sans motif consensus) déplacent l'ARNt marqué plus efficacement que les autres (Figure 41).

Nous avons choisi deux séquences : 26 et 23, chacune représentative des deux types d'aptamère (avec et sans séquence consensus, respectivement) pour vérifier leur capacité à dissocier le complexe *Pf*trip/ARNt$^{Asp}$ et à se fixer sur *Pf*trip et vérifier leur repliement en solution.


#30
#47
#17
#13
#28
#44
#32
#48
#14
#42
#46
#24
#15
#33
#27

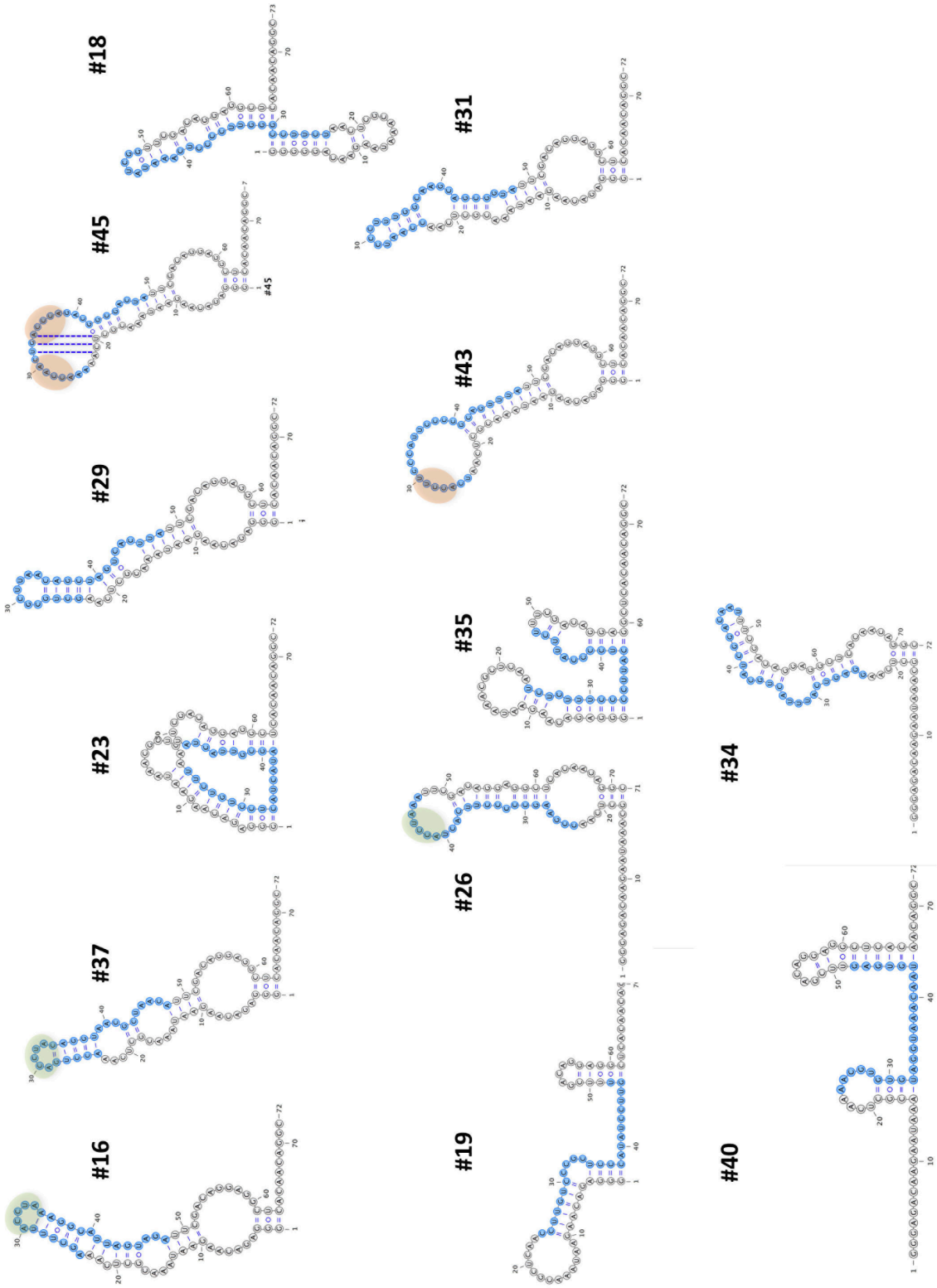



**Figure 40 : Prédiction de structures secondaires des aptamères sélectionnés.** Les repliements ont été prédits grâce à l'application RNAfold Webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) et les molécules ont été représentées (dans le même ordre que dans la Figure 39) grâce au logiciel Varna (VARNAv3-93). Les régions sélectionnées sont indiquées en bleu. Le motif 5'-ACCUA-3' est surligné en vert et en orange lorsque celui-ci contient une mutation.

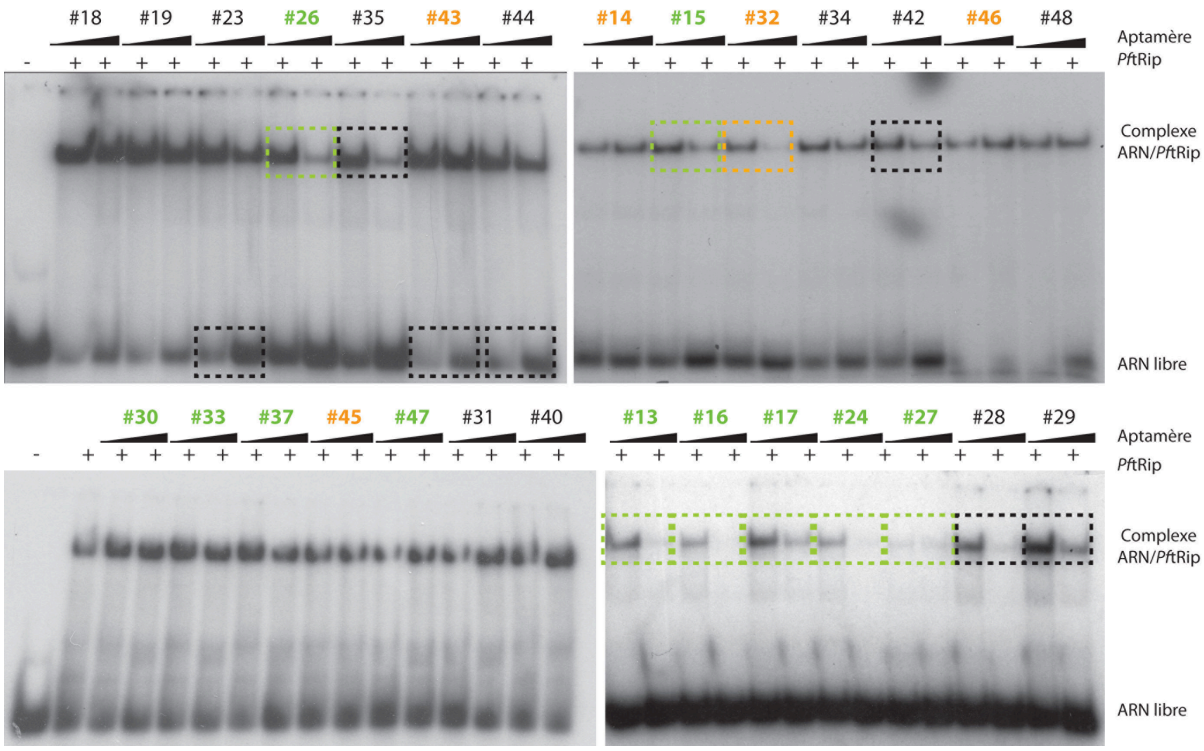



**Figure 41 : Criblage des différents aptamères pour leur capacité à se lier à *Pft*Rip**. Deux concentrations (40 et 160 nM) d'aptamères ont été utilisées pour tester leur capacité à dissocier le complexe préétabli entre *Pft*Rip et l'ARNt$^{Asp}$ de levure. Environ 100 nM d'ARNt total sont nécessaires pour observer une dissociation de ce complexe (pour un exemple, se reporter à la Figure 31 page 111).

### *2.5. Caractérisation des aptamères 23 et 26*

*a. Affinité pour PftRip*

Les expériences de compétition ont été refaites avec une gamme de concentrations d'aptamères plus étendue (de 0,44 à 3600 nM). L'ARN 23 commence à dissocier le complexe dès 30 nM alors que l'ARN 26, dissocie efficacement le complexe à une concentration plus élevée (225 nM) (Figure 42). Cette première observation a été confirmée par des expériences de retard sur gel « directs » où les aptamères 23 et 26 ont été marqués radioactivement et ont été incubés avec des concentrations croissantes de *Pft*Rip (de 31 à 500 nM) (Figure 43). Les deux ARN lient efficacement la protéine et conduisent à la formation d'oligomères de *Pft*Rip (comme le font les ARNt), cependant ces oligomères apparaissent plus rapidement avec l'ARN23 qu'avec l'ARN26, confirmant une meilleure efficacité de liaison de l'aptamère 23 à la protéine *Pft*Rip.

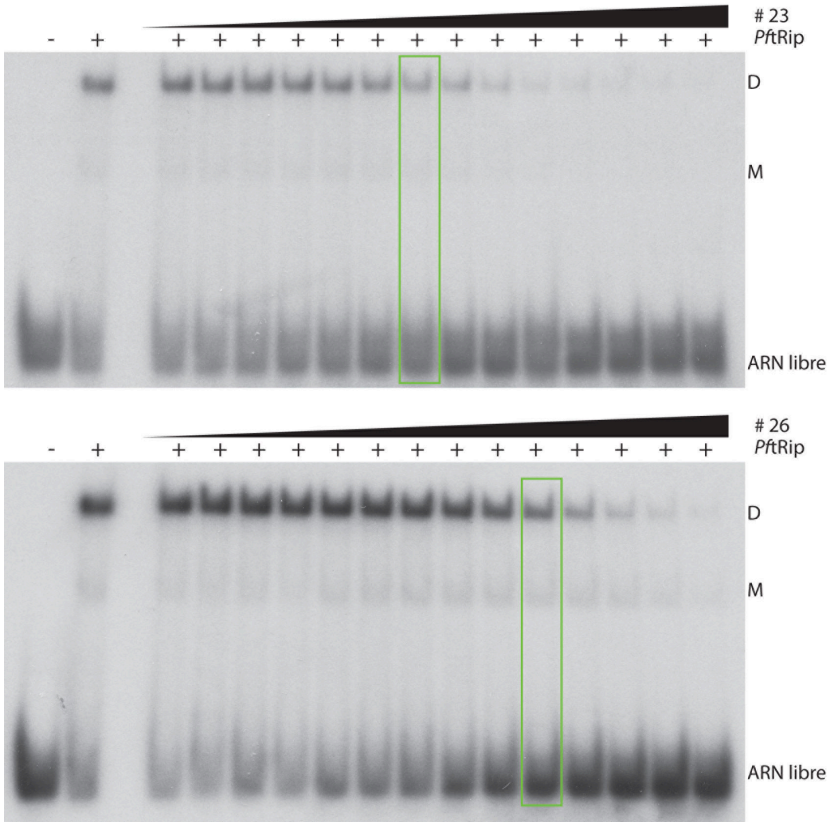



**Figure 42 : Capacité des aptamères 23 et 26 à se lier à *Pf*tRip.** Une gamme de concentrations d'aptamères allant de 0,44 à 3600 nM a été utilisée.

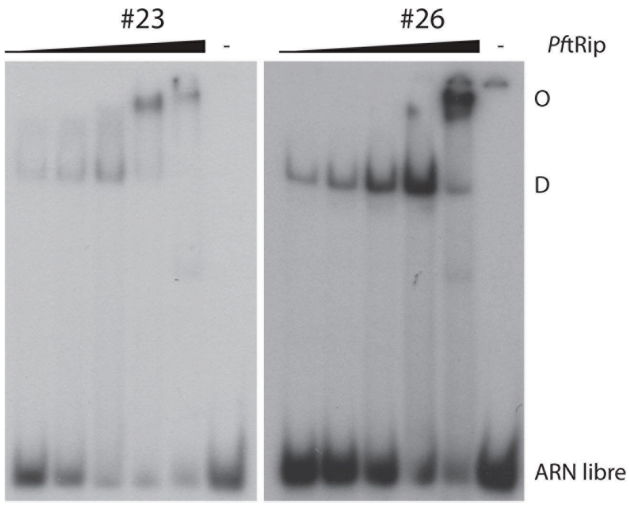



**Figure 43: Comparaison des capacités de liaison des aptamères 23 et 26 à *Pf*tRip par des expériences de retard sur gel « directs ».** Les ARN sont marqués radioactivements, leur concentration est fixe et limitante, alors que la concentration de *Pf*tRip augmente (31 nM, 62 nM, 125 nM, 250 nM et 500 nM). Comme dans un complexe formé entre *Pf*tRip et un ARNt, on observe deux complexes qui correspondent à un aptamère lié à un dimère de *Pf*tRip (D) et/ou un aptamère lié à un oligomère de *Pf*tRip de taille inconnu (O). Plus d'informations sur ces différents complexes sont donnée dans le chapitre consacré aux expériences de MIST (page 95).

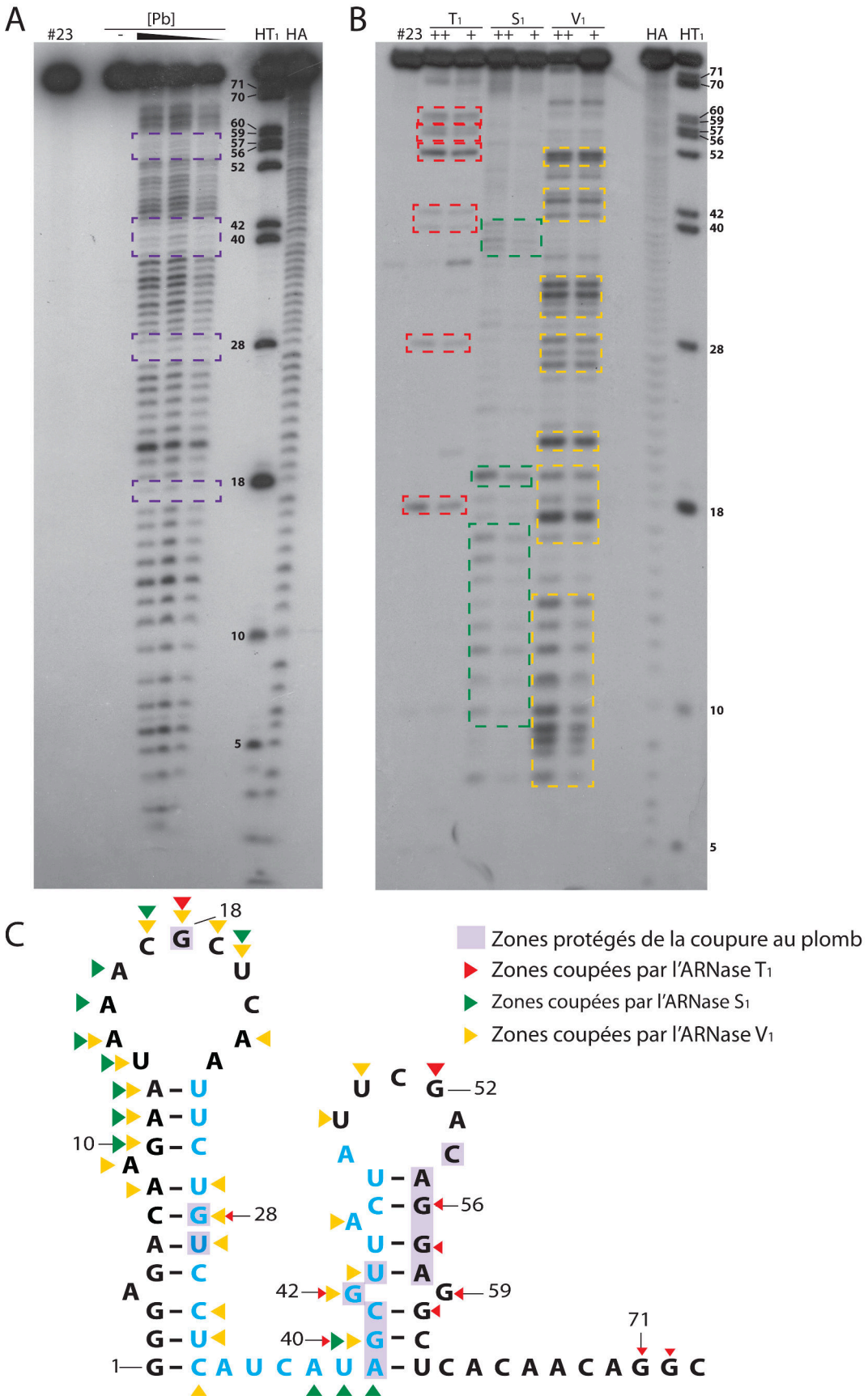

A
#23
[Pb]
-
HT1 HA
71
70
60
59
57
56
52
42
40
28
18
10
5
B
#23
T1
S1
V1
++ + ++ + ++ +
HA HT1
71
70
60
59
57
56
52
42
40
28
18
10
5
C
Zones protégés de la coupure au plomb
Zones coupées par l'ARNase T1
Zones coupées par l'ARNase S1
Zones coupées par l'ARNase V1
18
10
28
52
56
59
42
40
71
1
G – C A U C A U A – U C A C A A C A G G C

**Figure 44**

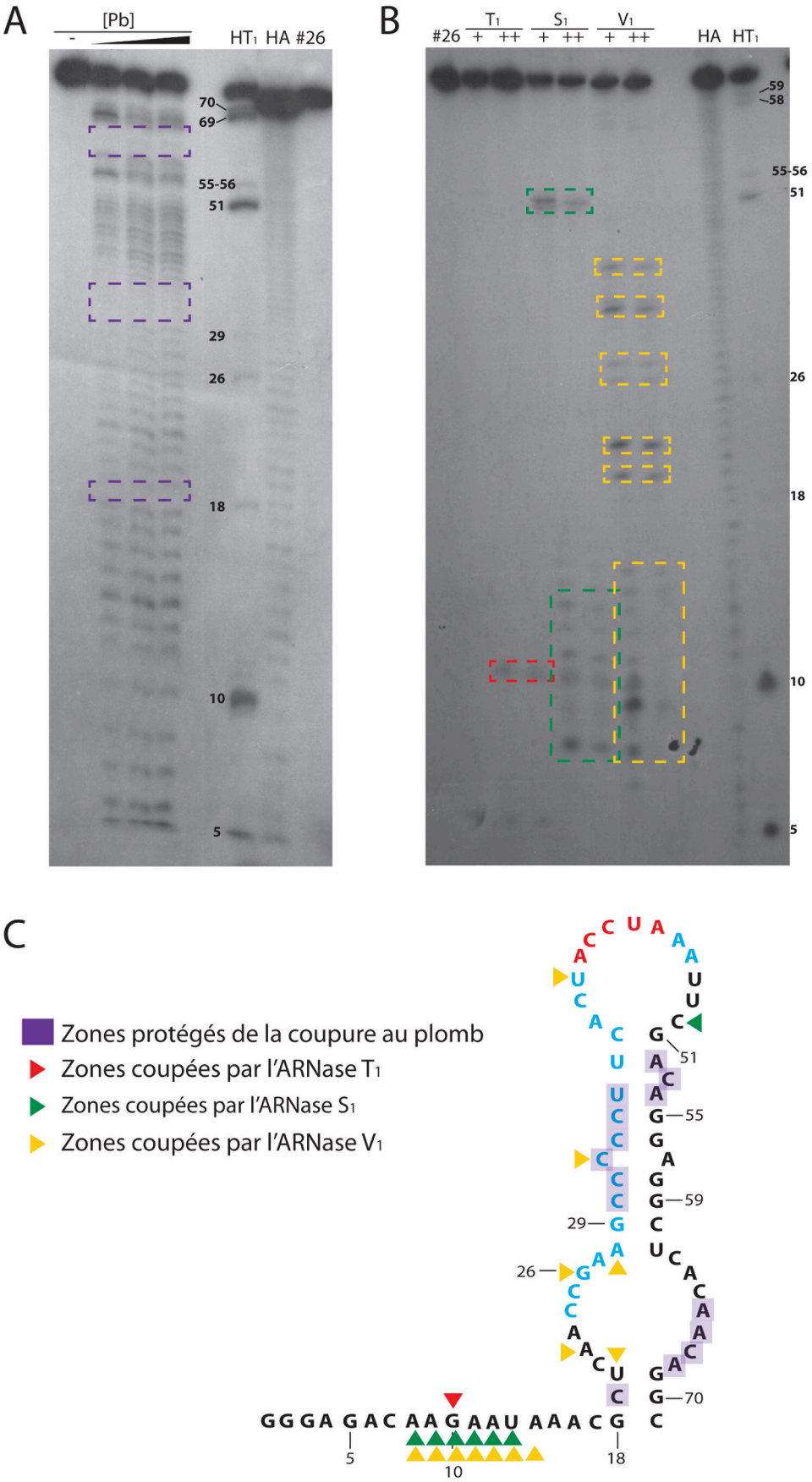

A
[Pb]
-
HT1 HA #26
70
69
55-56
51
29
26
18
10
5
B
#26
T1
S1
V1
+ ++
+ ++
+ ++
HA HT1
59
58
55-56
51
26
18
10
5
C
Zones protégés de la coupure au plomb
Zones coupées par l'ARNase T1
Zones coupées par l'ARNase S1
Zones coupées par l'ARNase V1
GGGA GAC AAGAAU AAACG
5
10
18
51
55
59
29
26
70

**Figure 45**

**Figure 44 : Etude de la structure de l'aptamère 23.** Autoradiographies **(A)** des coupures au plomb et **(B)** des enzymes T1, S1 et V1. La molécule d'ARN est marquée à son extrémité 5'. Le Pb coupe préférentiellement les zones non structurées, la RNAse T1 coupe après les résidus guanine localisés dans un simple brin, la RNase S1 coupe tous les nucléotides en simple brin et la RNase V1 coupe les nucléotides en double brin. **(C)** Récapitulatif des résultats du plomb et des enzymes T1, S1 et V1 sur la prédiction de structure de l'aptamère 23. Chaque expérience est accompagnée de trois contrôles : l'ARN incubé dans les mêmes conditions mais sans Pb ou RNAse (#23), une échelle T1 en conditions dénaturante (HT1) pour identifier les G et une échelle obtenue par hydrolyse alcaline (HA).

**Figure 45: Etude de la structure de l'aptamère 26.** Autoradiographie **(A)** des coupures du plomb et **(B)** des enzymes T1, S1 et V1. **(C)** Récapitulatif des résultats du plomb et des enzymes T1, S1 et V1 sur la prédiction de structure de l'aptamère 26. Pour plus de détail, voir la légende de la figure 44.

*b. Structures en solution*

Les structures secondaires des deux aptamères 23 et 26 ont été vérifiées en solution. Chaque ARN a été marqué radioactivement en 5 ', et incubé avec différentes concentrations de plomb (Pb) ou de RNAses (V1, T1 et S1). Les zones très structurées de l'ARN seront protégées de la coupure au plomb et coupées par la RNase V1 alors que les régions simple brin seront coupées par les RNases S1 et T1 (au niveau des résidus G) (Figures 44 et 45).

Les tests réalisés sur l'ARN 23 (Figure 44) montrent des résultats contradictoires où des zones de la région 5' sont protégées du Pb ne sont pas coupées par la RNase V1. De plus, ces mêmes zones sont coupées par la RNase S1. Ces observations indiquent que la structure de cet ARN n'est pas stable et que celui-ci adopte certainement des repliements alternatifs en solution.

Dans le cas de l'ARN 26 (Figure 45), les coupures sont beaucoup moins nombreuses, que ce soit dans les zones structurées ou les zones simple brin, ce qui pourrait s'expliquer par une structure de la molécule très compacte qui empêche l'accès aux RNAses et au Pb.

### *2.6. Conclusions*

Ces expériences de sélection de molécules d'ARN *in vitro* ont permis de sélectionner de nombreuses séquences qui reconnaissent spécifiquement le site de liaison aux ARNt de *Pf*tRip. Parmi ces molécules d'ARN, nous avons pu mettre en évidence un court motif nucléotidique qui pourrait être impliqué dans la spécificité de l'interaction. La molécule 26 porte ce motif et a fait l'objet d'une caractérisation plus poussée. Les premiers résultats

indiquent clairement que cet ARN se fixe sur *Pf*tRip, mais d'autres séquences devront être testées afin de d'identifier la ou les molécules qui ont le plus d'affinité pour *Pf*tRip. Il sera alors nécessaire de confirmer leur repliement secondaire *in vitro* et compléter ces résultats par des expériences d'empreintes de *Pf*tRip sur l'aptamère. Cette approche devrait nous permettre de mettre en évidence l'implication ou pas du motif consensus dans la reconnaissance des aptamères par *Pf*tRip. Enfin la conception d'ARN plus courts et de mutants sera essentielle pour déterminer si ce motif est suffisant pour une reconnaissance optimale par *Pf*tRip.

## III. PRESENCE DES ARNT DE L'HOTE DANS LE PARASITE AU STADE SANGUIN

Dans l'étude initiale par Bour et coll. (Bour, *et al.*, 2016), l'import des ARNt exogènes dans le parasite a été démontré au stade sporozoïte uniquement. Deux approches ont été utilisées pour faire cette démonstration :

(1) les sporozoïtes purifiés à partir de glandes salivaires de moustiques infectés ont été incubés en présence d'ARNt$^{Val}$ d'*E. coli* pendant 15 min à température ambiante, lavés et utilisés pour faire une expérience de FISH (Fluorescence in situ hybridization) grâce à une sonde fluorescente complémentaire et spécifique de l'ARNt$^{Val}$ d'*E. coli*. La détection d'un signal fluorescent dans le sporozoïte indique que l'import a eu lieu.

(2) des sporozoïtes, débarrassés des contaminants cellulaires provenant des glandes salivaires de moustiques ont été incubés avec de l'ARNt total de cellules HeLa marqué radioactivement pendant 15 min à température ambiante, lavés, puis traités à la RNAse A (pour éliminer tous les ARNt qui n'ont pas été importés). Après un phénol, les ARN des sporozoites ont été analysés sur gel dénaturant. La présence d'une bande radioactive migrant à la même distance qu'un ARNt indique que l'import d'un ARNt entier a eu lieu dans le parasite.

Chacune des deux expérience a été accompagnée d'un contrôle où le sporozoïte a été préalablement atténué par la chaleur (30 min à 50°C, ce traitement tue le parasite mais ne détruit pas son intégrité). Dans les deux cas, le sporozoïte atténué n'importe pas d'ARNt.

Dans ce même article, il a également été montré qu'au stade sanguin, *Pb*tRip est localisée à la surface du parasite, qu'en l'absence de *Pb*tRip (parasite *Pb*tRip-KO) le développement du parasite est ralenti et que la synthèse protéique est réduite. Cependant, la démonstration de l'import d'ARNt dans le stade sanguin n'a pas été faite. L'objectif de mon travail a donc été de rechercher la présence d'ARNt de l'hôte dans le stade sanguin du parasite du paludisme malgré des conditions très défavorables.

Les conditions défavorables dont il est question sont résumées dans la Figure 46. Contrairement au sporozoïte pendant sa traversée cellulaire, le stade sanguin est caractérisé par la présence d'une vacuole parasitophore qui l'isole du cytosol du globule rouge. De plus, l'hyper spécialisation du globule rouge mature, qui est une cellule anucléée et sans traduction active, correspond à une situation où on peut se demander si il y a des ARNt à importer. Finalement, les séquences des ARNt du parasite sont très similaires aux séquences

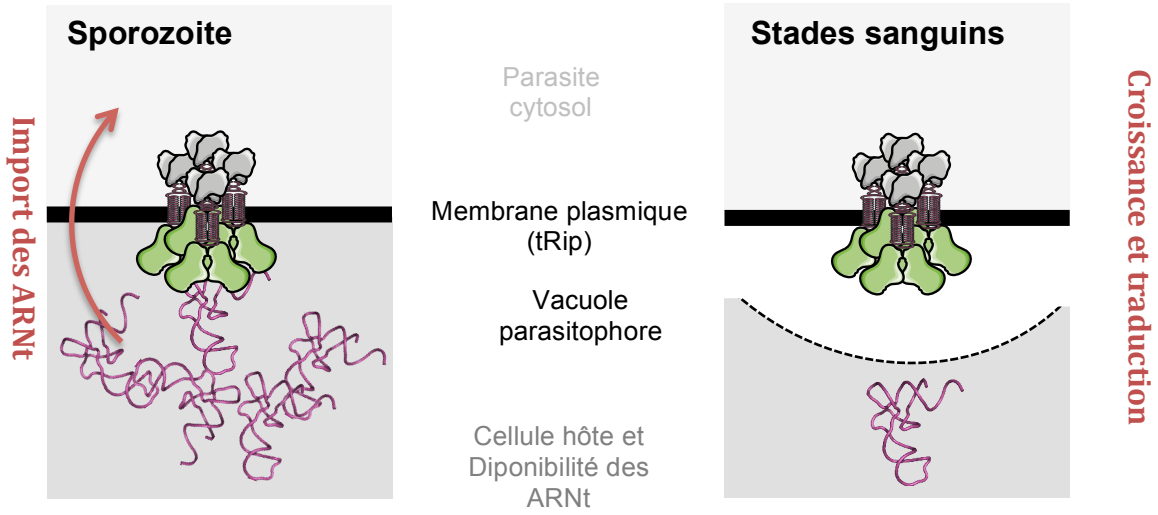



**Figure 46 : Représentation de tRip dans le contexte des membranes de *Plasmodium* et de la disponibilité des ARNt de l'hôte**. Nous avons démontré que les sporozoïtes importent des ARNt exogènes et que les parasites tRip-KO se développent plus lentement avec une synthèse protéique réduite par rapport aux parasites de type sauvage. tRip est ancrée dans la membrane plasmique du parasite, le domaine N-terminal (gris) étant à l'intérieur du parasite et le domaine C-terminal (vert) étant à l'extérieur au contact des ARNt de l'hôte.

des ARNt de vertébrés, ce qui rend la conception de sondes spécifiques capables de différencier les ARNt de souris des ARNt des *Plasmodium* particulièrement difficile.

Il m'a donc fallu répondre à plusieurs questions :

### *1. Y a t-il de l'ARNt dans les globules rouges ?*

Afin de mettre en évidence la présence ou pas d'ARNt dans les globules rouges (GR), j'ai sacrifié des souris naïves adultes. Le sang a, dans un premier temps, été débarrassé de tous les globules blancs en utilisant un filtre plasmodipur. La disparition des cellules nucléées été vérifiée sous microscope. A partir de 2,5 ml de sang (soit environ 2500 mg de sang) j'ai obtenu 8 µg d'ARN total. Ce rendement est particulièrement faible puisque à partir de 420 mg de foie de souris, j'obtiens 1250 µg d'ARN total.

Sur le gel de la Figure 47A, on observe que les ARN purifiés à partir de foie ou de sang de souris ont le même profil. Cependant le rendement de la purification indique que les

globules rouges contiennent environ 100 fois moins d'ARN que des cellules de foie (cellules hôte des sporozoïtes).

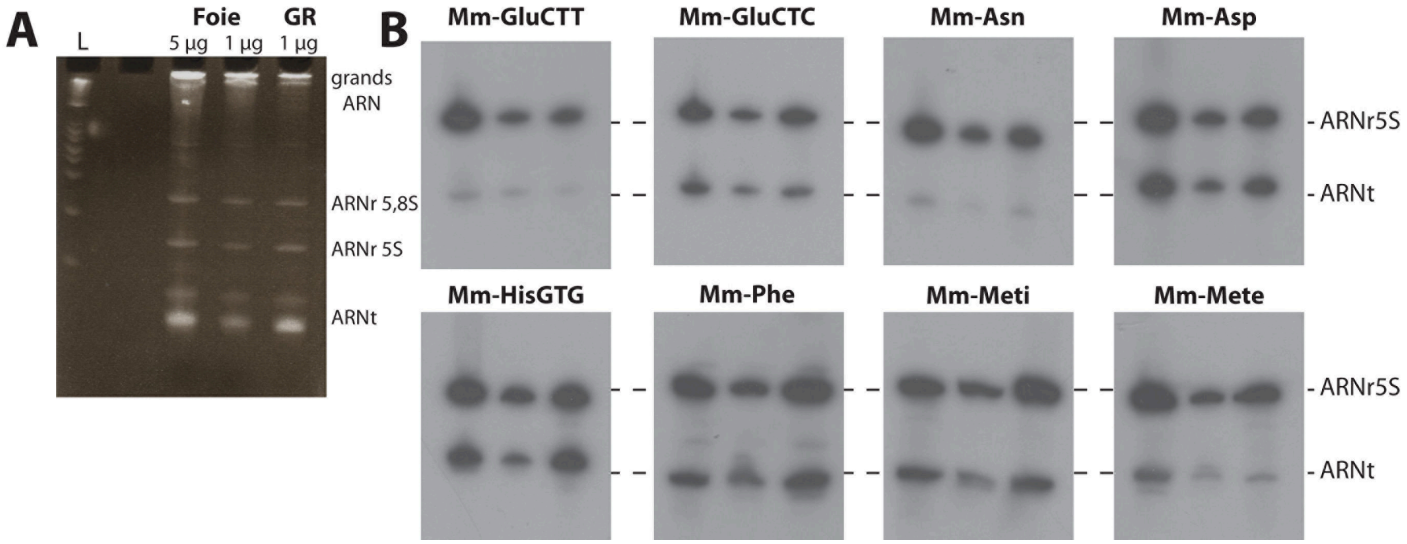



**Figure 47: Comparaison des ARN isolés à partir de foie ou de globules rouges (GR) de souris. (A)** Gel analytique dénaturant 12 % coloré au BET. 5 et 1 µg d'ARN total de foie et 1 µg d'ARN total de GR sont comparés. L correspond à une échelle de poids moléculaires d'ADN. **(B)** Northern blot après hybridation avec des sondes conçues pour détecter spécifiquement les ARNt de souris Mm- GluTTC, GluCTC, Asn, Asp, HisGTG, Phe, $Met_i$ et $Met_e$. L'hybridation de la sonde dirigée contre l'ARN ribosomique 5S (ARNr5S) de souris permet de vérifier la quantité d'ARN déposée dans chaque piste.

Huit sondes dirigées spécifiquement contre les ARNt $Glu_{CTC}$, $Glu_{TTC}$, Asp, Asn, $His_{GTG}$, Phe, Meti et Mete de souris ont été conçues (Figure Supplémentaire 3) et testées sur les échantillons d'ARN extraits de foie ou de GR de souris par Northern blot (Figure 47B). En se basant sur l'intensité de la bande correspondant à l'ARN ribosomique 5S (ARNr5S), on peut conclure que l'ARN de GR contient globalement la même proportion d'ARNt que l'ARN de cellules de foie.

## *2. Mise au point d'un protocole de purification des ARN de parasite*

Au cours de ces expériences, j'ai été confrontée à une difficulté majeure. D'une part, la majorité des souris sont efficacement infectées par le parasite sauvage (WT), mais les souris sont malades ou meurent alors que la parasitémie est encore faible et dépasse rarement 5 à

7 %. D'autre part, peu de souris sont infectées par le parasite KO, le parasite se développe lentement, mais lorsqu'elles sont infectées, la parasitémie atteint des pourcentages élevés (parfois jusqu'à 20 %, sans que les souris aient de symptômes). Ensemble, ces contraintes expérimentales font qu'il est difficile d'obtenir suffisamment de sang infecté par les parasites WT et KO, pour effectuer des purifications d'ARN en parallèle. La plupart du temps les purifications d'ARN de parasites ont été faites séparément, et les différents lots d'ARN récupérés ont été cumulés pour faire les Northerns blot présentés dans ce chapitre.

J'ai développé un protocole qui permet l'extraction de l'ARN total du parasite, en minimisant la contamination externe de l'ARN de la souris. Ce protocole consiste en quatre étapes : (1) l'élimination du sérum et de la majorité des globules blancs par centrifugation, (2) la séparation des GR infectés des GR non-infectés sur un gradient de percoll, (3) une lyse des membranes des érythrocytes pour libérer les parasites et (4) une étape finale où les parasites sont incubés avec de la RNase A pour éliminer les ARN de souris qui restent associés à la surface du parasite.

### *2.1. Effet de la RNase*

Sur le gel de la Figure 48, cinq purifications d'ARN de parasite KO sont comparées. Elles diffèrent par les temps d'incubation en présence de RNases A qui varient entre 5 et 20 min à température ambiante. Dans ces conditions, on voit que les ARNr 5,8 et 5S ainsi que les ARNt sont épargnés par la dégradation à la RNase alors que les grands ARN du parasite ne sont plus visibles dès 15 min d'incubation avec la RNase A. Bien que ces conditions soient drastiques, puisque les ARN de grandes tailles sont dégradés, nous avons choisi d'incuber les parasites 15 min à température ambiante pour être sûrs que les ARNt de souris contaminants potentiellement accrochés sur le parasite sont complètement dégradés par la RNase.

### *2.2. Rendements*

En comparant les rendements de plusieurs purifications, on constate que l'efficacité varie considérablement d'une expérience à l'autre, qu'elle est faible et qu'elle a tendance à diminuer en même temps que la parasitémie initiale (Tableau 9). Les meilleures purifications permettent au maximum d'obtenir 2 µg d'ARN de parasite par souris infectée.

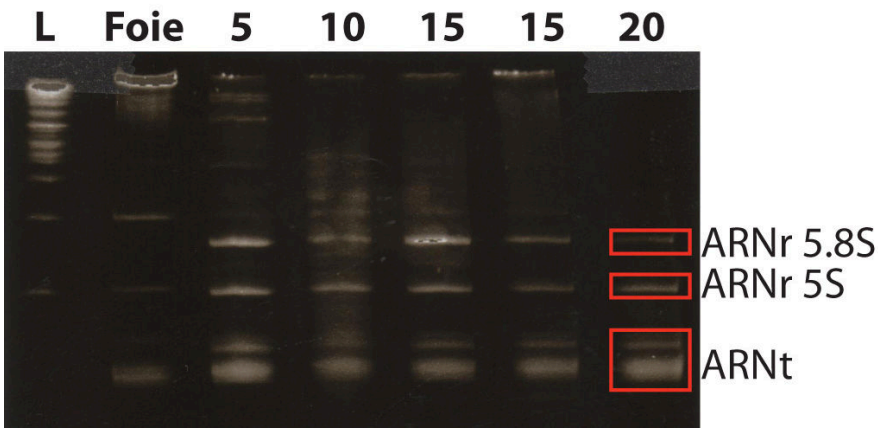



**Figure 48 : Comparaison de différentes préparations d'ARN isolés à partir de parasites *Pb*tRip KO.** La dernière étape de purification d'ARN de *Plasmodium* au stade sanguin consiste en la dégradation des ARN de souris contaminants par de la RNase A. Les 5 échantillons de parasites ont été incubés 5, 10, 15 ou 20 minutes en présence de 2,5 µg de RNase A à température ambiante. L correspond à une échelle de poids moléculaires. Un échantillon d'ARN de foie de souris a été déposé comme contrôle.

**Tableau 9 : Efficacités de purifications des ARN de *Plasmodium berghei***

| Valeurs réelles | | | Valeurs théoriques | | Efficacités |
|---|---|---|---|---|---|
| Parasitémie | µL de sang | µg d'ARN purifié | GRi | µg ARN | (%) |
| 10,6 | 1000 | 1,06 | 848000 | 2,54 | 41,7 |
| 6,9 | 1000 | 0,89 | 552000 | 1,66 | 53,6 |
| 7,5 | 1000 | 0,14 | 600000 | 1,8 | 7,8 |
| 6,7 | 1000 | 0,14 | 536000 | 1,61 | 8,7 |
| 3,5 | 700 | 0,05 | 196000 | 0,59 | 8,5 |
| 8,3 | 1000 | 0,03 | 664000 | 2 | 1,5 |

Six purifications ont été effectuées à partir de sang de souris infecté par le parasite WT (*Pb* GOMO14). Les valeurs indiquées correspondent à la parasitémie (en %), le volume de sang utilisé pour la purification (correspond à chaque fois à une souris sacrifiée) et à la quantité d'ARN total récupéré. Les valeurs théoriques ont été calculées à partir d'une estimation du nombre de GR dans 1 mL de sang de souris (8. $10^6$ GR/ml) et de la quantité d'ARN total dans un parasite (3 pg). Pour ces calculs, on a considéré que chaque globule rouge infecté (GRi) contient un parasite au maximum.

### *2.3. Qualité des échantillons*

Deux échantillons d'ARN de parasite WT et KO ont été testés par Northern blot. Afin de déterminer le niveau de contamination des préparations par de l'ARN de souris, des sondes conçues pour détecter les ARNr5S de souris et de parasite (contrôle de charge) et spécifiquement l'ARN 7SL et l'ARNt$^{Sec}$ de souris ont été utilisées (Figure Supplémentaire 3). L'ARN 7SL de souris, n'étant pas importé par le parasite, il constitue un bon témoin pour mettre en valeur la pureté de nos échantillons. Cette expérience (et de nombreuses autres)

a permis de mettre en évidence que les ARN de souris sont bien dégradés par le traitement à la RNase A, mais pas suffisamment pour obtenir des échantillons exploitables. En effet, La sonde Mm-7SL détecte une seule bande dans l'ARN de foie de souris, mais aussi dans les échantillons d'ARN de parasites où le signal est diffus et correspond à un ARN partiellement dégradé (Figures 49A et 49B). La sonde dirigée contre l'ARNt$^{Sec}$ de souris détecte aussi un faible signal dans les ARN de parasite, en particulier dans l'ARN de parasite KO, confirmant que ce protocole de purification ne permet pas d'éliminer tous les contaminants.

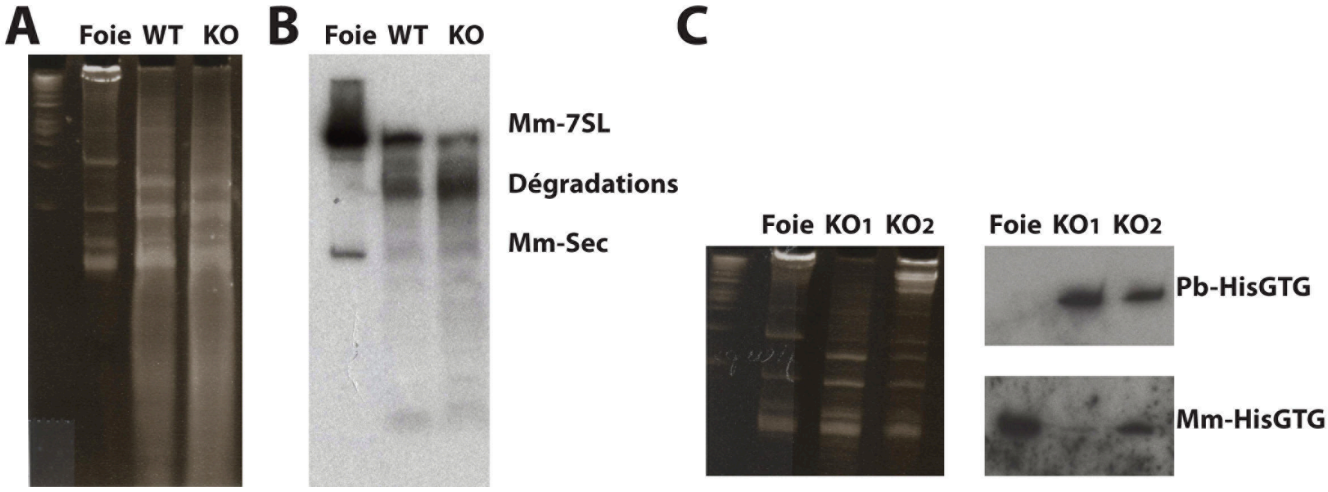



**Figure 49: Qualité des préparations d'ARN de parasites après traitement à la RNase A. (A)** Gel analytique dénaturant 12 % coloré au BET. 1 µg d'ARN total de foie de souris et 1 µg d'ARN total de parasite WT et KO sont analysés. L correspond à une échelle de poids moléculaires. **(B)** Northern blot correspondant au gel (A) après hybridation avec des sondes conçues pour détecter spécifiquement l'ARN7SL et l'ARNt$^{Sec}$ de souris. **(C)** Gel analytique et Northern blot. 1 µg d'ARN total de foie de souris et 1 µg d'ARN total de deux préparations indépendantes d'ARN de parasite KO (KO1 et KO2) sont analysés. L'hybridation de la sonde dirigée contre l'ARNt$^{His}_{GTG}$ de *P. berghei* (Pb-HisGTG) permet de vérifier l'origine des ARN étudiés alors que l'hybridation de la sonde dirigée contre l'ARNt$^{His}_{GTG}$ de souris (Mm-HisGTG) permet d'estimer le niveau de contamination de chaque échantillon.

De même, sur deux échantillons d'ARN de parasite KO (KO1 et KO2) qui ont a priori été purifiés selon le même protocole (Figure 49C), on voit qu'ils contiennent globalement la même quantité d'ARNt$^{His}$ de parasite (Pb-His) mais sont tout de même contaminés à des degrés très différents par de l'ARNt de souris (Mm-His). On est conscient que la faible reproductibilité de ces expériences ne nous permet pas d'utiliser ce protocole pour tester la présence d'ARNt de souris dans le parasite aux stades sanguins.

### *3. Peut-on éviter l'utilisation de RNases ?*

Dans un deuxième temps, les Northern blot ont été faits sur des échantillons d'ARN obtenus à partir de sang infecté, débarrassé des globules blancs sur filtre plasmodipur et directement utilisé pour l'extraction de l'ARN.

Le gel de la Figure 50A montre que les profils des échantillons issus de parasite WT ou KO sont similaires, mais différents du profil de migration de l'ARN de souris. Chez les parasites, on distingue les grand ARN qui correspondent en particulier aux ARN ribosomaux 28 et 18S (3790 et 2150 nucléotides, respectivement) aux ARNr 5.8S et 5S (168 et 119 nucléotides, respectivement) et les ARNt. Alors que chez la souris, seuls les petits ARNr 5.8S et 5S (156 et 121 nucléotides) sont bien visibles. En revanche on ne distingue pas les grands ARNr28 et 18S (4712 et 1870 nucléotides, respectivement).

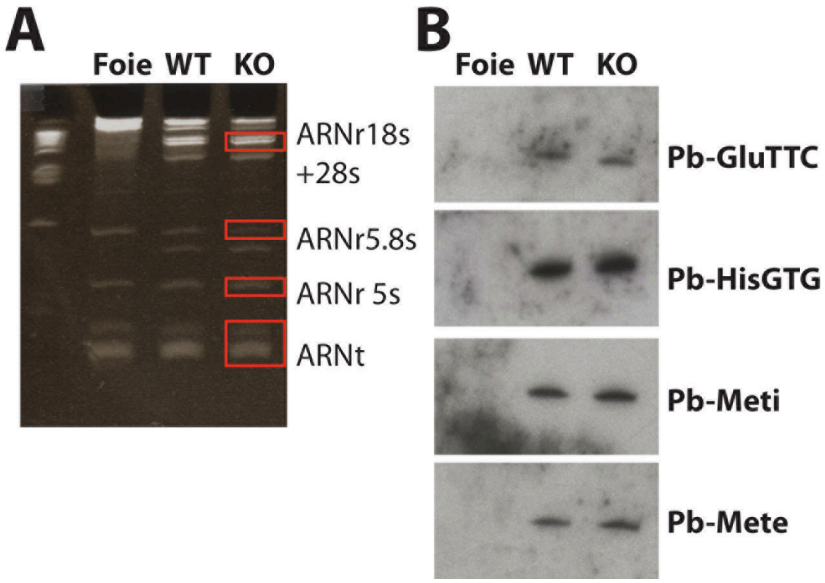



**Figure 50 : Qualité des préparations d'ARN de parasites sans traitement à la RNase. (A)** Gel analytique dénaturant 12 % coloré au BET. 1 µg d'ARN total de foie de souris et 1 µg d'ARN total de parasite WT et KO sont analysés. L correspond à une échelle de poids moléculaires. **(B)** Northern blot correspondant au gel (A) après hybridation avec des sondes conçues pour détecter spécifiquement les $ARNt^{Glu}_{TTC}$, $His_{GTG}$, $Met_i$ et $Met_e$ de *P. berghei* (Pb-GluTTC, Pb-HisGTG, Pb-Meti et Pb-Mete).

Ces échantillons d'ARN ont été utilisés dans des expériences de Northern blot afin de déterminer si les sondes conçues pour détecter les ARNt de *Plasmodium* sont efficaces et spécifiques. Pour cela 4 sondes conçues pour s'hybrider aux $ARNt^{Glu}_{TTC}$, His, $Met_i$ et $Met_e$ (Figure Supplémentaire 3) ont été testées (Figure 50B). Les sondes s'hybrident

essentiellement aux ARNt présents dans les échantillons d'ARN purifiés à partir des parasites et pas à l'ARN de souris, indiquant qu'elles sont effectivement spécifiques.

### *4. Peut-on détecter des ARNt de souris parmi les ARNt du parasite ?*

Quatre membranes ont été préparées avec de l'ARN de foie de souris et de l'ARN de parasite WT et KO et testées avec les sondes conçues pour détecter spécifiquement les $ARNt^{Glu}_{TTC}$, His, $Met^{i}$ et $Met_e$ de souris (Figure 51). La première expérience de Northern blot sur ces membranes a mis en évidence des signaux plus ou moins intenses dans tous les échantillons, même dans l'ARN de parasite KO alors qu'il n'est pas capable d'importer les ARNt de l'hôte (Figure 51A). Ce premier résultat suggère que les sondes ne sont pas suffisamment spécifiques ou que les ARN de parasites ne sont pas suffisamment propres et contiennent encore des ARN de souris.

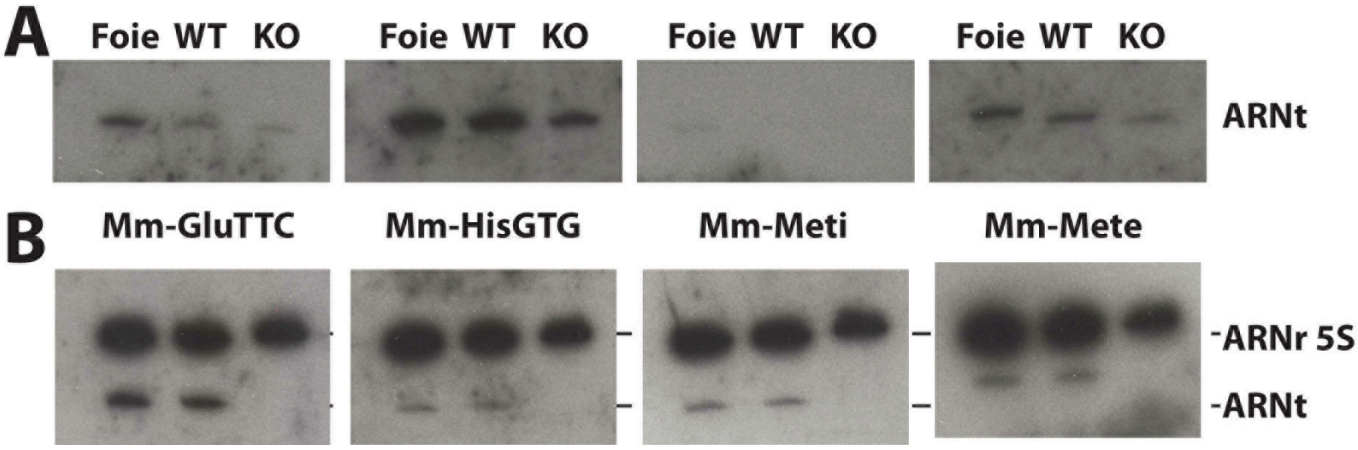



**Figure 51 : Détection des ARNt de souris dans le parasite par Northern blots.** Les mêmes membranes ont été hybridées avec **(A)** des sondes dirigées spécifiquement contre les $ARNt^{Glu}_{TTC}$, $His_{GTG}$, $Met_i$ et $Met_e$ de souris (Mm-GluTTC, MmHisGTG, Mm-Meti, Mm-Mete) et **(B)** ces mêmes sondes et la sonde dirigée contre l'ARN ribosomique 5S (ARNr5S qui s'hybride aux ARNr5S de souris et du parasite). 1 µg d'ARN total de foie de souris et 1 µg d'ARN total de parasite WT et KO sont analysés.

Cependant, les mêmes membranes ont été réhybridées avec la sonde dirigée contre les deux ARN5S de souris et de *Plasmodium*. De façon intéressante, après une exposition prolongée, les sondes dirigées contre les ARNt $Glu_{TTC}$, $His_{GTG}$, $Met_i$ et $Met_e$ indiquent clairement que de l'ARNt de souris est présent dans le parasite WT mais pas dans le parasite

KO, alors que les contrôles de charge (ARNr5S) montrent que les dépôts d'ARN sont équivalents pour tous les échantillons (Figure 51B). Ce résultat suggère que des lavages plus importants permettent d'éliminer les sondes dirigées contre les ARNt de souris qui se sont hybridées de façon non spécifique aux ARNt du parasite.

### *5. Conclusions*

Il est possible de détecter spécifiquement des ARNt de l'hôte dans les stades sanguins du parasite WT, mais pas dans le parasite KO. Cependant, ces expériences sont difficilement reproductibles à cause des hybridations non spécifiques entre les sondes de souris et les ARNt de *Plasmodium*. Nous avons pu mettre en évidence qu'il n'est pas judicieux d'utiliser de la RNase A pour éliminer les ARN de souris qui contaminent l'extérieur du parasite puisque celle-ci dégrade aussi partiellement les ARN endogènes du parasite. De façon surprenante les filtres plasmodipur semblent suffisamment efficaces pour éliminer la majeure partie des ARN de souris, certainement en éliminant les cellules autres que les GR dès le début du protocole de purification. Enfin, la plus grande limite de ces expériences est le rendement extrêmement faible des purifications d'ARN du parasite.

Une autre approche que les Northern blot pourrait nous permettre de confirmer ces premières données et d'atteindre notre but. Nous voulons utiliser la technique de protection aux RNAses (A et T1 qui coupent essentiellement les nucléotides non appariés). Dans ce cas, il s'agit d'hybrider en solution les ARNt du parasite à des sondes ARN radioactives et spécifiques et de digérer tout ce qui n'est pas complètement apparié. Ainsi, il est possible de discriminer 2 séquences d'ARN qui diffèrent seulement par 1 nucléotide (Dietrich, *et al*., 1996).

D'autre part, il nous reste à connaître l'origine des ARNt importés dans les parasites au stade sanguin. Bien qu'il ait été possible de purifier des ARNt à partir de globules rouges, le rendement de cette purification montre que comme attendu, la quantité d'ARNt dans ces cellules est faible. Cependant, le sang contient aussi des réticulocytes, qui ne sont pas éliminés par les filtres plasmodipur. De façon intéressante, il a été montré que malgré la rareté des réticulocytes dans le flux sanguin (environ 2% des cellules), *Plasmodium berghei* envahit préférentiellement ces cellules (150 fois plus) par rapport aux GR matures (Cromer, *et al*., 2006. Or ces réticulocytes, qui sont des précurseurs de GR, sont très actifs en

traduction et contiennent beaucoup d'ARNt (Smith et McNamara, 1972). Si nous voulons savoir si les stades sanguins de *Plasmodium* trouvent préférentiellement leur source d'ARNt dans les réticulocytes, il est possible d'augmenter artificiellement la quantité de réticulocytes dans une souris en lui injectant de l'érythropoïétine (EPO) et de vérifier si dans ces conditions, le développement du parasite WT est favorisé alors que le développement du parasite KO ne devrait pas être affecté.

## FIGURES SUPPLEMENTAIRES

**Ala1h**

TGGAGAATGYGGGCGTCGATCCCRCTACCTCTYGCATGCTAAGCRAGCGCTCTACCRCCTGAGCTAATTCCCC

tRNAAlaIGC

```
>>>>>>>..>>.>.........<.<<.>>>>>.......<<<<<.....>>>>>.......<<<<<<<<<<<....
GGGGAATTAGCTCAAATGGTA.GAGCGCTCGCTTAGCATGCGAGAGGtAGCGGGATCGATGCCCGCATTCTCCA...tRNA-Ala-AGC-11-1
GGGGAATTAGCTCAAGTGGTA.GAGCGCTTGCTTAGCATGCAAGAGGtAGTGGGATCAATGCCCACATTCTCCA...tRNA-Ala-AGC-14-1
GGGGAATTAGCTCAAGCGGTA.GAGCGCTTGCTTAGCATGCAAGAGGtAGTGGGATCGATGCCCACATTCTCCA...tRNA-Ala-AGC-13-1
GGGGAATTAGCTCAAGCGGTA.GAGCGCTTGCTTAGCATGCAAGAGGtAGTGGGATCGATGCCCACATTCTCCA...tRNA-Ala-AGC-13-2
GGGGAATTAGCTCAGGCGGTA.GAGCGCTCGCTTAGCATGCGAGAGGtAGCGGGATCGACGCCCGCATTCTCCA...tRNA-Ala-AGC-9-1
GGGGAATTAGCTCAGGCGGTA.GAGCGCTCGCTTAGCATGCGAGAGGtAGCGGGATCGACGCCCGCATTCTCCA...tRNA-Ala-AGC-9-2
GGGGAATTAGCTCAAGTGGTA.GAGCGCTCGCTTAGCATGCGAGAGGtAGTGGGATCGATGCCCGCATTCTCCA...tRNA-Ala-AGC-15-1
GGGGAATTAGCCCAAGTGGTA.GAGCGCTTGCTTAGCATGCAAGAGGtAGTGGGATCGATGCCCACATTCTCCA...tRNA-Ala-AGC-16-1
```

**Ala2h**

TGGAGRTGCCGGGGATYGAACCCGGGRCCTCRTRCATGCSAAGCAYGCGCTCTACCACTGAGCTACAYCCCC

tRNAAlahGC

```
>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<.....>.>>>.........<<<.<<<<<<<<....
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCGTGCTTAGCATGCACGAGGcCCCGGG.TTCAATC.CCCGGCACCTCCA...tRNA-Ala-AGC-2-2
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCGTGCTTAGCATGCACGAGGcCCCGGG.TTCAATC.CCCGGCACCTCCA...tRNA-Ala-AGC-2-1
GGGGGTATAGCTCAGT.GGT.A..GAGCGCGTGCTTAGCATGCACGAGGtCCTGGG.TTCGATC.CCCAGTACCTCCA...tRNA-Ala-AGC-1-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCGTGCTTAGCATGCACGAGGcCCTGGG.TTCAATC.CCCAGCACCTCCA...tRNA-Ala-AGC-5-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCGTGCTTAGCATGTACGAGGtCCCGGG.TTCAATC.CCCGGCACCTCCA...tRNA-Ala-AGC-3-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCGTGCTTAGCATGCACGAGGcCCCGGG.TTCAATC.CCTGGCACCTCCA...tRNA-Ala-AGC-7-1

GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTCGCATGTATGAGGtCCCGGG.TTCGATC.CCCGGCATCTCCA...tRNA-Ala-CGC-1-1
GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTCGCATGTATGAGGcCCCGGG.TTCGATC.CCCGGCATCTCCA...tRNA-Ala-CGC-2-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCGTGCTTCGCATGTACGAGGcCCCGGG.TTCGACC.CCCGGCTCCTCCA...tRNA-Ala-CGC-4-1
GGGGATGTAGCTCAGT.GGT.A..GAGCGCGCGCTTCGCATGTGTGAGGtCCCGGG.TTCAATC.CCCGGCATCTCCA...tRNA-Ala-CGC-3-1
GGGGGTGTAGATCAGT.GGT.A..GAGCGCATGCTTCGCATGTACGAGGtCCCTGG.TTCAATC.CCTGGTACCTCCA...tRNA-Ala-CGC-5-1

GGGGGTGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGtCCCGGG.TTCGATC.CCCGGCACCTCCA...tRNA-Ala-TGC-1-1
GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGtCCCGGG.TTCGATC.CCCGGCATCTCCA...tRNA-Ala-TGC-2-1
GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGcCCCGGG.TTCGATC.CCCGGCATCTCCA...tRNA-Ala-TGC-3-2
GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGcCCCGGG.TTCGATC.CCCGGCATCTCCA...tRNA-Ala-TGC-3-1
GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCACGTATGAGGcCCCGGG.TTCAATC.CCCGGCATCTCCA...tRNA-Ala-TGC-4-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGcCTCGGG.TTCGATC.CCCGACACCTCCA...tRNA-Ala-TGC-5-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCACATGCTTTGCATGTGTGAGGcCCCGGG.TTCGATC.CCCGGCACCTCCA...tRNA-Ala-TGC-6-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGG.CCTCGG.TTCGATC.CCCGACACCTCCA...tRNA-Ala-TGC-7-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCGGATGCTTTGCATGTATGAGAcTTTGGG.TTGGATC.CCCAGCACCTCCA...tRNA-Ala-TGC-8-1
GTGGATGTAGTTTAGT.GGT.A..GAACGCGCGCTTTGCATGTATGAGGtCCCGGT.TTCGATC.CCTGGCGTTTCCA...tRNA-Ala-TGC-9-1
```

**Arg1h**

CGAGCCAGCCAGGAGTCGAACCTRGAATCTTCTGATCCGTAGTCAGACGCGTTATCCATTGCGCCACTGGCCC

tRNAArgiCG

```
.>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<.<<<<<<<.
.GGGCCAGTGGCGCAATGGAt.A.ACGCGTCTGACTACGGATCAGAAGAtTCCAGGTTCGACTCCTGG.CTGGCTCG...tRNA-Arg-ACG-1-3
.GGGCCAGTGGCGCAATGGAt.A.ACGCGTCTGACTACGGATCAGAAGAtTCCAGGTTCGACTCCTGG.CTGGCTCG...tRNA-Arg-ACG-1-2
.GGGCCAGTGGCGCAATGGAt.A.ACGCGTCTGACTACGGATCAGAAGAtTCCAGGTTCGACTCCTGG.CTGGCTCG...tRNA-Arg-ACG-1-1
.GGGCCAGTGGCGCAATGGAt.A.ACGCGTCTGACTACGGATCAGAAGAtTCTAGGTTCGACTCCTGG.CTGGCTCG...tRNA-Arg-ACG-2-4
.GGGCCAGTGGCGCAATGGAt.A.ACGCGTCTGACTACGGATCAGAAGAtTCTAGGTTCGACTCCTGG.CTGGCTCG...tRNA-Arg-ACG-2-3
.GGGCCAGTGGCGCAATGGAt.A.ACGCGTCTGACTACGGATCAGAAGAtTCTAGGTTCGACTCCTGG.CTGGCTCG...tRNA-Arg-ACG-2-2
.GGGCCAGTGGCGCAATGGAt.A.ACGCGTCTGACTACGGATCAGAAGAtTCTAGGTTCGACTCCTGG.CTGGCTCG...tRNA-Arg-ACG-2-1
.GGGCCAGTGGCGCAATGGAt.A.ACGCGTCTGACTACGGATCAGAAGAtTCCAGGTTCGACTCCTGG.CTGGCTCG...DR9990_ACG_HUMAN_CY
```

**Arg2h**

CAACCACGRAGGGACTCGAACCCTCAATCTTCTGATCCGRARTCAGACGCCTTATCCATTAGGCCACGCGGCC

tRNAArgyCG

```
.>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<.<<<<<<<.
.GGCCGCGTGGCCTAATGGAt.A.AGGCGTCTGATTCCGG.ATCAGAAG.AtTGAGGGTTCGAGTCCCTT.CGTGGTCG...tRNA-Arg-CCG-1-3
.GGCCGCGTGGCCTAATGGAt.A.AGGCGTCTGATTCCGG.ATCAGAAG.AtTGAGGGTTCGAGTCCCTT.CGTGGTCG...tRNA-Arg-CCG-1-2
.GGCCGCGTGGCCTAATGGAt.A.AGGCGTCTGATTCCGG.ATCAGAAG.AtTGAGGGTTCGAGTCCCTT.CGTGGTCG...tRNA-Arg-CCG-1-1
.GACCCAGTGGCCTAATGGAt.A.AGGCATCAGCCTCCGG.AGCTGGGG.AtTGTGGGTTCGAGTCCCAT.CTGGGTCG...tRNA-Arg-CCG-2-1
.GGCCGCGTGGCCTAATGGAt.A.AGGCGTCTGACTTCGG.ATCAGAAG.AtTGCAGGTTCGAGTCCTGC.CGCGGTCG...tRNA-Arg-TCG-1-1
.GACCACGTGGCCTAATGGAt.A.AGGCGTCTGACTTCGG.ATCAGAAG.AtTGAGGGTTCGAATCCCTC.CGTGGTTA...tRNA-Arg-TCG-2-1
.GACCACGTGGCCTAATGGAt.A.AGGCGTCTGACTTCGG.ATCAGAAG.AtTGAGGGTTCGAATCCCTT.CGTGGTTA...tRNA-Arg-TCG-4-1
.GACCGCGTGGCCTAATGGAt.A.AGGCGTCTGACTTCGG.ATCAGAAG.AtTGAGGGTTCGAGTCCCTT.CGTGGTCG...tRNA-Arg-TCG-3-1
.GACCACGTGGCCTAATGGAt.A.AGGCGTCTGACTTCGG.ATCAGAAG.AtTGAGGGTTCGAATCCCTT.CGTGGTTG...tRNA-Arg-TCG-5-1
.GACCACGTGGCCTAATGGAt.A.AGGCGTCTGACTTCGG.ATCAGAAG.AtTGAGGGTTCGAATCCCTT.CGTGGTT. DR9991_TCG_HUMAN_CY
```

**Arg3h**

TACCCCAGGTGGGACTCGAACCCACAATCCCTGGCTTAGGAGGCCAGTGCCTTATCCATTAGGCCACTGGGGC

tRNAArgCCT

```
.>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<.<<<<<<<.
.GCCCCAGTGGCCTAATGGAt.A.AGGCACTGGCCTCCTA.AGCCAGGG.AtTGTGGGTTCGAGTCCCAC.CTGGGGTA tRNA-Arg-CCT-1-1
.GCCCCAGTGGCCTAATGGAt.A.AGGCACTGGCCTCCTA.AGCCAGGG.AtTGTGGGTTCGAGTCCCAC.CTGGGGTG tRNA-Arg-CCT-2-1
.GCCCCGGTGGCCTAATGGAt.A.AGGCATTGGCCTCCTA.AGCCAGGG.AtTGTGGGTTCGAGTCCCAC.CCGGGGTA tRNA-Arg-CCT-3-1
.GCCCCAGTGGCCTAATGGAt.A.AGGCATTGGCCTCCTA.AGCCAGGG.AtTGTGGGTTCGAGTCCCAT.CTGGGGTG tRNA-Arg-CCT-4-1
.GCCCCAGTGGCCTGATGGAt.A.AGGTACTGGCCTCCTA.AGCCAGGG.AtTGTGGGTTCGAGTTCCAC.CTGGGGTA tRNA-Arg-CCT-5-1
```

**Arg4h**

CGACTCTGCYGGGACTCGAACCCGWAACCTTTGAATTAGAAGTCCAATGCGCTATCCATTGCGCCACAGAGCC

tRNAArgTCT

```
.>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<.....>.>>>.......<<<.<.<<<<<<<.
.GTCTCTGTGGCGCAATGGAcgA.GCGCGCTGGACTTCTAATCCAGAGGtTCCGGGTTCGAGTCCCGG.CAGAGATG tRNA-Arg-TCT-4-1
.GGCTCTGTGGCGCAATGGAt.A.GCGCATTGGACTTCTAATTCAAAGGtTGTGGGTTCGAATCCCAC.CAGAGTCG tRNA-Arg-TCT-2-1
.GGCTCCGTGGCGCAATGGAt.A.GCGCATTGGACTTCTAATTCAAAGGtTCCGGGTTCGAGTCCCGG.CGGAGTCG tRNA-Arg-TCT-1-1
```

```
.GGCTCTGTGGCGCAATGGAt.A.GCGCATTGGACTTCTAATTCAAAGGtTGTGGGTTCGAGTCCCAC.CAGAGTCG tRNA-Arg-TCT-3-2
.GGCTCTGTGGCGCAATGGAt.A.GCGCATTGGACTTCTAATTCAAAGGtTGTGGGTTCGAGTCCCAC.CAGAGTCG tRNA-Arg-TCT-3-1
.GGCTCTGTGGCGCAATGGAt.A.GCGCATTGGACTTCTAATTCAAAGGtTGCGGGTTCGAGTCCCTC.CAGAGTCG tRNA-Arg-TCT-5-1
```

## Asn1h

CGTCCCTGGGTGGGCTCGAACCACCAACCTTTCGGTTAACAGCCGAACGCGCTAACCGATTGCGCCACAGAGAC

### tRNAAsnGTT

```
>>>>>>>..>>>>..........<<<<.>>>>>.........<<<<<.....>>>>>.......<<<<<<<<<<<....
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGATCCCACCCAGGGACG... tRNA-Asn-GTT-1-1
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-2-6
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-2-5
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-2-4
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-2-3
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-2-2
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-2-1
GTCTCTGTGGCGCAATCGGTtA.GCGCATTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-3-2
GTCTCTGTGGCGCAATCGGTtA.GCGCATTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-3-1
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGAtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-4-1
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACTGAAAGGtTGGTGGTTCGAGCCCACCCAGGGACG... tRNA-Asn-GTT-5-1
GTCTCTGTGGCGCAATGGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCATCCAGGGACG... tRNA-Asn-GTT-6-1
GTCTCTGTGGCGCAATCGGCtA.GCGCGTTTGGCTGTT..AACTAAAAGGtTGGTGGTTCGAACCCACCCAGAGGCG... tRNA-Asn-GTT-8-1
GTCTCTGTGGCGTAGTCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAAGtTGGTGGTTCGAGCCCACCCAGGAACG... tRNA-Asn-GTT-7-1
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACTGAAAGGtTAGTGGTTCGAGCCCACCCGGGGACG... tRNA-Asn-GTT-9-2
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACTGAAAGGtTAGTGGTTCGAGCCCACCCGGGGACG... tRNA-Asn-GTT-9-1
GTCTCTGTGGCGCAATCGGCtA.GCGCGTTTGGCTGTT..AACTAAAAGGtTGGCGGTTCGAACCCACCCAGAGGCG... tRNA-Asn-GTT-10-1
GTCTCTGTGGTGCAATCGGTtA.GCGCGTTCCGCTGTT..AACCGAAAGCtTGGTGGTTCGAGCCCACCCAGGGATG... tRNA-Asn-GTT-11-2
GTCTCTGTGGTGCAATCGGTtA.GCGCGTTCCGCTGTT..AACCGAAAGCtTGGTGGTTCGAGCCCACCCAGGGATG... tRNA-Asn-GTT-11-1
GTCTCTGTGGCGCAATCGGCtA.GCGCGTTTGGCTGTT..AACTAAAAAGtTGGTGGTTCGAACACACCCAGAGGCG... tRNA-Asn-GTT-12-1
GTCTCTGTGGTGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCATAAGGtTGGTGGTTAGAGACCACCCAGGGACG... tRNA-Asn-GTT-13-1
GTCTCTGTGGTGCAATCGGTtA.GCGCGTTCCGCTGTT..AACCGAAAGCtTGGTGGTTTGAGCCCACCCAGGGATG... tRNA-Asn-GTT-14-1
GTCTCTGTGGCGCCATCGGTtA.GTGCCTTCGGCTGTTtgAACCGAAAGGcTGGTGGTTCAAGCCCACCCAGAGATG... tRNA-Asn-GTT-20-1
GTTTCTGTAGCGCGATCGGTtA.GCGCCTTCGGCTGTT..AAACGAAAGGtTGGTGGTTCGTTCCCACCCCGGGACA... tRNA-Asn-GTT-17-1
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTTGACTGTT..AACTGAAAGGtTGGTGGTGCAAGCCCATCCAGGGATG... tRNA-Asn-GTT-16-5
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTTGACTGTT..AACTGAAAGGtTGGTGGTGCAAGCCCATCCAGGGATG... tRNA-Asn-GTT-16-4
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTTGACTGTT..AACTGAAAGGtTGGTGGTGCAAGCCCATCCAGGGATG... tRNA-Asn-GTT-16-3
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTTGACTGTT..AACTGAAAGGtTGGTGGTGCAAGCCCATCCAGGGATG... tRNA-Asn-GTT-16-2
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTTGACTGTT..AACTGAAAGGtTGGTGGTGCAAGCCCATCCAGGGATG... tRNA-Asn-GTT-16-1
GTCTCTGTGGTGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCATAAGGtTGGTGGTTACAGACCACCCAGGGACG... tRNA-Asn-GTT-18-1
ATCTCCGTGGAGCAATTGGTtA.GCGCGTTCGGCTGTT..AACCGGAAAGtTGGTGGTTCGAGCCTACCCAGGGACG... tRNA-Asn-GTT-15-1
ATCTCCGTGGAGCAATTGGTtA.GCGCGTTCGGCCGTT..AACCGGAAAGtTGGTGGTTCGAGCCTACCCAGGGACG... tRNA-Asn-GTT-19-1
GTCTCTGTGGCACAATCGGTtA.GCGCGTTCGGCTGTT..AATCTAGAGGtTGGTGGTTAGAGCCCACTGAGGGATG... tRNA-Asn-GTT-21-1
GTCTCTGTGGCACAATCGGTtA.GAGCGTTCGGCTGTT..AATCTAAAGGtTGGTGGCTAGAACCCACTGAGGGACG... tRNA-Asn-GTT-22-1
GTCTCTGTGGCGCAATGGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGTTCGAGCCCATCCAGGGACG... DN9990_GTT_HUMAN_CY
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGAtTGGTGGTTCGAGCCCACCCAGGGACG... DN9991_GTT_HUMAN_CY
GTCTCTGTGGCGCAATCGGTtA.GCGCGTTCGGCTGTT..AACCGAAAGGtTGGTGGNTCGAGCCCACCCAGGGACGcca RN9990_GTT_HUMAN_LIVER/PLAC
```

## Asp1h

CTCCCCGTCGGGGAATCGAACCCCGGTCTCCCGCGTGACAGGCGGGGATACTCACCACTATACTAACGAGGA

### tRNAAspGTC

```
>>>>>>>..>>>>.........<<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<.<<<<<<<.
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-9
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-8
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-7
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-6
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-5
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-4
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-3
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-2
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-1
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-11
TCCTCGTTAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-2-10
TCCTCGTTAGTATAGTGGTtA.GTATCCCCGCCTGTCACGCGGGAGA.CCGGGGTTCAATTCCCCG.ACGGGGAG tRNA-Asp-GTC-1-1
TCCTCGTTAGTATAGTGGTgA.GTGTCCCCGTCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCG.ACGGGGAG tRNA-Asp-GTC-3-1
TCCTCGTTAGTATGGTGGTgA.GTATCCCTGCCTGTCACGCGGGAGA.CCGGGGTTCGATTCCCCA.ACGGGGAG tRNA-Asp-GTC-4-1
TCCTCATCAGTATAGTGGTgA.GTATCCCCGCCTGTCACGCGGGAGA.CTGGGGTTCGATTCCCTG.AGGAGGAG tRNA-Asp-GTC-5-1
TTCTTGTTAATATAGTGGTgA.GTATTCCCACCTGTCATGCGGGA.G.ACGGGGTTCAATTCCCTG.ATGGGGAG tRNA-Asp-GTC-6-1
TCCTTGTTAGTATAGTGGTgA.GTGTTTCTGCCTGTCATGTGGAG.A.CTGGAGTTTGAGTCCCCA.ACAGGGAG tRNA-Asp-GTC-8-1
```

## Glu1h

TTCCCTGACCGGGAATCGAACCCGGGCCWMGGCGGTGARAGCGCCGAATCCTARCCACTAGACCACCAGGGR

### tRNAGluyTC

```
>>>>>>>..>>>>.........<<<<.>>>>>.......<<<<.<....>>>>>.......<<<<<..<<<<<<<....
TCCCTGGTGGTCTAGTGGTtA.GGATTCGGCGCTCTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-CTC-1-7
TCCCTGGTGGTCTAGTGGTtA.GGATTCGGCGCTCTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-CTC-1-6
TCCCTGGTGGTCTAGTGGTtA.GGATTCGGCGCTCTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-CTC-1-5
TCCCTGGTGGTCTAGTGGTtA.GGATTCGGCGCTGTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-CTC-1-4
TCCCTGGTGGTCTAGTGGTtA.GGATTCGGCGCTCTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-CTC-1-3
TCCCTGGTGGTCTAGTGGTtA.GGATTCGGCGCTCTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-CTC-1-2
TCCCTGGTGGTCTAGTGGTtA.GGATTCGGCGCTCTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-CTC-1-1
TCCCTGGTGGTCTAGTGGTtA.GGATTCGGCGCTCTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGAAA... tRNA-Glu-CTC-2-1
CCCCTGGTGGTCTAGTGCTtA.GGATTCGGTGCTCTCACCGCT.GCTGCCTGCGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-CTC-3-1
TCCTTGAT.GTCTAGTGGTtA.GGATTTGGTGCTCTCACTGCA.GCAGCCTGGGTTCATTTCTCAG..TCAGGGAA... tRNA-Glu-CTC-5-1
TCCCTGGTGGTCTAGTGGTtA.GGATTCGGCGCTCTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAAcca RE9990_CTC_HUMAN_CY

TCCCTGGTGGTCTAGTGGCtA.GGATTCGGCGCTTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..CCAGGGAA... tRNA-Glu-TTC-3-1
TCCCTGGTGGTCTAGTGGCtA.GGATTCGGCGCTTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-TTC-4-2
TCCCTGGTGGTCTAGTGGCtA.GGATTCGGCGCTTCACCGCC.GCGGCCCGGGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-TTC-4-1
TCCCTGGTGGTCTAGTGGCtA.GGATTCGGCGCTTCACCGCCtGCAGCTCGAGTTCGATTCCTGG..TCAGGGAA... tRNA-Glu-TTC-8-1
TCCTTGGTGGTCTAGTGGCtA.GGATTCGGTGCTTCACCTGT.GCGGCCCGGGTTCAATTCCCGA..TGAAGGAA... tRNA-Glu-TTC-6-1
TCCGTGGTGGTCTAGTGGCtA.GGATTCGGCGCTTCACCGCCtGCAGCTCGAGTTCGATTCCTGG..TCAGGGAA... tRNA-Glu-TTC-10-1
..CCCTGTGGTCTAGTGGCtA.AGACTTTGTGCTTTCATTGCT.GC.ACCTAGGTTCAATTCCCAG..T.CAGGGA...tRNA-Glu-TTC-11-1
GCAATGGTGGTTCAGTGGT.A.GAATTCTCGCCTTCACACAG.GAGACCCGGGTTCAATTCCTGA..CCCATGTA... tRNA-Glu-TTC-9-1
TGTCTGGTGGTCAAGTGGCtA.GGATTTGGCGCTTCACTGCC.GCGGCCCGCGTTCGATTCCCGG..TCAGGGAA... tRNA-Glu-TTC-7-1
GCGTTGGTGGTGTAGTGGTgA.GCACAGCTGCCTTTCAAGCAG.TTAACGCGGGTTCGATTCCCGG..GTAACGAA... tRNA-Glu-TTC-5-1
```

## Glu2h

TTCCCAYACCGGGAGTCGAACCCGGGCCRCCTGGGTGAAAACCAGGAATCCTAACCGCTAGACCATRTGGGA

### tRNAGluTTC-1

```
>>>>>>>..>>>>.........<<<<.>>>>>.......<<<<.<....>>>>>.......<<<<<..<<<<<<<....
TCCCACATGGTCTAGCGGTtA.GGATTCCTGGTTTTCACCCAG.GCGGCCCGGGTTCGACTCCCGG..TGTGGGAA... tRNA-Glu-TTC-2-2
TCCCACATGGTCTAGCGGTtA.GGATTCCTGGTTTTCACCCAG.GCGGCCCGGGTTCGACTCCCGG..TGTGGGAA... tRNA-Glu-TTC-2-1
TCCCATATGGTCTAGCGGTtA.GGATTCCTGGTTTTCACCCAG.GTGGCCCGGGTTCGACTCCCGG..TATGGGAA... tRNA-Glu-TTC-1-2
TCCCATATGGTCTAGCGGTtA.GGATTCCTGGTTTTCACCCAG.GTGGCCCGGGTTCGACTCCCGG..TATGGGAA... tRNA-Glu-TTC-1-1
```

## Glu3h

TTCGTTACCCGGGAATCGAACCCGCGTTAACTGCTTGAAAGGCAGCTGTGCTCACCACTACACCACCAACGC

### tRNAGluTTC-2

```
>>>>>>>..>>.>.........<.<<.>>>>>.......<<<<<.....>>>>.......<<<<.<<<<<<<....
GCGTTGGTGGTGTAGTGGTGAGCACAGCTGCCTTTCAAGCAGTTAACGCGGGTTCGAT.TCCCGGGTAACGAA... tRNA-Glu-TTC-5-1
```

### tRNAGlyTCC

```
>>>>>>>..>>.>.........<.<<.>>>>>.......<<<<<.....>>>>.......<<<<.<<<<<<<....
GCGTTGGTGGTATAGTGGTtA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-1-1
GCGTTGGTGGTATAGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-6
GCGTTGGTGGTATAGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-5
GCGTTGGTGGTATAGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-4
GCGTTGGTGGTATAGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-3
GCGTTGGTGGTATAGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-2
GCGTTGGTGGTATAGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-1
GCGTTGGTGGTATAGTGGTaA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-3-1
GCGTTGGTGGTATAGTGGTgA.GCATAGTTGCCTTCCAAGCAGTTGACCCGGGCTCGATTCCCGCCCAACGCA...tRNA-Gly-TCC-4-1
```

## Gly1h

TGCATTGGCCRGGAATYGAACCCGGGYCTCCCRCGTGGWAGGCGAGAATTCTACCACTGMACCACCMAYGC

### tRNAGlysCC

```
>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<....>>>>>.......<<<<<<<<<<<<....
GCATGGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCCATGCA... tRNA-Gly-GCC-1-5
GCATGGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCCATGCA... tRNA-Gly-GCC-1-4
GCATGGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCCATGCA... tRNA-Gly-GCC-1-3
GCATGGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCCATGCA... tRNA-Gly-GCC-1-2
GCATGGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCCATGCA... tRNA-Gly-GCC-1-1
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCAATGCA... tRNA-Gly-GCC-2-6
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCAATGCA... tRNA-Gly-GCC-2-5
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCAATGCA... tRNA-Gly-GCC-2-4
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCAATGCA... tRNA-Gly-GCC-2-3
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCAATGCA... tRNA-Gly-GCC-2-2
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCAATGCA... tRNA-Gly-GCC-2-1
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTTGATTCCCGGCCAGTGCA... tRNA-Gly-GCC-3-1
GCATAGGTGGTTC.AGTGGT.A.GAATTCTTGCCTGCCACGCAGGAGGCCCAGGTTTGATTCCTGGCCCATGCA... tRNA-Gly-GCC-4-1
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCATGCGGGCGGCCGGGCTTCGATTCCTGGCCAATGCA... tRNA-Gly-GCC-5-1
GCATGGGTGATTC.AGTGGT.A.GAATTTTCACCTGCCATGCAGGAGGTCCAGGTTCATTTCCTGGCCTATGCA... tRNA-Gly-GCC-6-1
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCAATGCA... DG9990_GCC_HUMAN_CY
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCAATGCAcca RG9991_GCC_HUMAN_PLACENTA

GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTCCCACGCGGGAGACCCGGGTTCAATTCCCGGCCAATGCA... tRNA-Gly-CCC-1-2
GCATTGGTGGTTC.AGTGGT.A.GAATTCTCGCCTCCCACGCGGGAGACCCGGGTTCAATTCCCGGCCAATGCA... tRNA-Gly-CCC-1-1
GCGCCGCTGGTGT.AGTGGT.A.TCATGCAAGATTCCCATTCTTGCGACCCGGGTTCGATTCCCGGGCGGCGCA... tRNA-Gly-CCC-2-2
GCGCCGCTGGTGT.AGTGGT.A.TCATGCAAGATTCCCATTCTTGCGACCCGGGTTCGATTCCCGGGCGGCGCA... tRNA-Gly-CCC-2-1
GCATTGGTGGTTC.AATGGT.A.GAATTCTCGCCTCCCACGCAGGAGACCCAGGTTCGATTCCTGGCCAATGCA... tRNA-Gly-CCC-3-1
GCGTTGGTGGTTT.AGTGGT.A.GAATTCTCGCCTCCCATGCGGGAGACCCGGGTTCAATTCCCGGCCACTGCA... tRNA-Gly-CCC-4-1
GCCTTGGTGGTGC.AGTGGT.A.GAATTCTCGCCTCCCACGTGGGAGACCCGGGTTCAATTCCCGGCCAATGCA... tRNA-Gly-CCC-5-1
GCGTCAGTGGTTT.AGTGGT.G.GAATTCCTGCCTCCCATGCACGAGATCCGTGTTCAACTCCTGGTTGGTGCA... tRNA-Gly-CCC-8-1
GCGTTGGCAGTTC.AGTGGT.A.GAATTCTCGCCTCCCACCCGGGAGACCTGGATTCCATTTCCGGCAAATGCA... tRNA-Gly-CCC-chr1-19
GCGTTGGCAGTTC.AGTGGT.A.GAATTCTCGCCTCCCACCCGGGAGACCTGGATTCCATTTCCGGCAAATGCA... tRNA-Gly-CCC-chr1-16
GCGTCAGTGGTTTtAGTGGT.G.GAATTCCTGCCTCCCATGCACGAGATCCGTGTTCAACTCCTGGTTGGTGCA... tRNA-Gly-CCC-chr1-102
```

## Gly2h

TGCGTTGGCCGGGAATCGAACCCGGGTCAACTGCTTGGAAGGCAGCTATGCTCACCACTATACCACCAACGC

### tRNAGlyTCC

```
>>>>>>>..>>.>.........<.<<.>>>>>.......<<<<<....>>>>>.......<<<<<<<<<<<<....
GCGTTGGTGGTAT.AGTGGTtA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-1-1
GCGTTGGTGGTAT.AGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-6
GCGTTGGTGGTAT.AGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-5
GCGTTGGTGGTAT.AGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-4
GCGTTGGTGGTAT.AGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-3
GCGTTGGTGGTAT.AGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-2
GCGTTGGTGGTAT.AGTGGTgA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-2-1
GCGTTGGTGGTAT.AGTGGTaA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA...tRNA-Gly-TCC-3-1
GCGTTGGTGGTAT.AGTGGTgA.GCATAGTTGCCTTCCAAGCAGTTGACCCGGGCTCGATTCCCGCCCAACGCA...tRNA-Gly-TCC-4-1
GCGTTGGTGGTAT.AGTGGTtA.GCATAGCTGCCTTCCAAGCAGTTGACCCGGGTTCGATTCCCGGCCAACGCA... DG9992_TCC_HUMAN_CY
```

## His1h

TGCCGTGACTCGGATTCGAACCGAGGTTGCTGCGGCCACAACGCAGAGTACTAACCACTATACGATCACGGC

### tRNAHisGTG

```
.>>>>>>>..>>>>.........<<<<.>>>>>.......<<<<<....>>>>>.......<<<<<<<<<<<<....
.GCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-1-9
.GCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-1-8
.GCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-1-7
.GCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-1-6
.GCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-1-5
.GCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-1-4
.GCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-1-3
.GCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-1-2
.GCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-1-1
.GCCATGATCGTATAGTGGTtA.GTACTCTGCGCTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA...tRNA-His-GTG-2-1
gGCCGTGATCGTATAGTGGTtA.GTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCAcca RH9990_GTG_HUMAN_HELA-CELLS
```

**Ile2h**

TGCTCCAGGTGAGGMTCGAACTCACAACCTCGGCATTATAAGTACCGCGCGCTAACCGATTGCGCCACTGGAGC

tRNAIleTAT

```
>>>>>>>...>>>>.........<<<<.>>>>>.......<<<.<<.....>>>>>.......<<<<<<<<<<<<.
GCTCCAGT.GGCGCAATCGGTtAGCGCGCGGTACTTATAATGC.CGAGGtTGTGAGTTCGATCCTCACCTGGAGCA...tRNA-Ile-TAT-1-1
GCTCCAGT.GGCGCAATCGGTtAGCGCGCGGTACTTATAATGC.CGAGGtTGTGAGTTCGAGCCTCACCTGGAGCA...tRNA-Ile-TAT-2-2
GCTCCAGT.GGCGCAATCGGTtAGCGCGCGGTACTTATAATGC.CGAGGtTGTGAGTTCGAGCCTCACCTGGAGCA...tRNA-Ile-TAT-2-1
GCTCCAGT.GGCGCAATCGGTtAGCGCGCGGTACTTATAATGC.CGAGGtTGTGAGTTCGAGCCTCACCTGGAGCA...tRNA-Ile-TAT-2-3
GCTCCAGT.GGCGCAATCGGTtAGCGCGCGGTACTTATAATGC.CGAGGtTGTGAGTTCAAGCCTCACCTGGAGCA...tRNA-Ile-TAT-3-1
```

**Leu1h**...

TGGCAGYGGTGGGATTCRAACCCACGCCYCCGAARAGACTGGAGCCTTAATCCAGCGCCTTAGACCRCTCGGCCACRCTA

tRNALeuwAG

```
>>>>>>>..>>>.............<<<.>>>>>........<<<<<...>>>>...........<<<<...>>>>>.......<<<<<.<<<<<<<....
GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTAAGG.CTCCAGT..CTCT..TC........GGAGG.CGTGGGTTCGAATCCCAC.CGCTGCCA...tRNA-Leu-AAG-1-1
GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTAAGG.CTCCAGT..CTCT..TC........GGAGG.CGTGGGTTCGAATCCCAC.CGCTGCCA...tRNA-Leu-AAG-1-2
GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTAAGG.CTCCAGT..CTCT..TC........GGAGG.CGTGGGTTCGAATCCCAC.CGCTGCCA...tRNA-Leu-AAG-1-3
GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTAAGG.CTCCAGT..CTCT..TC........GGGGG.CGTGGGTTCGAATCCCAC.CGCTGCCA...tRNA-Leu-AAG-2-1
GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTAAGG.CTCCAGT..CTCT..TC........GGGGG.CGTGGGTTCGAATCCCAC.CGCTGCCA...tRNA-Leu-AAG-2-2
GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTAAGG.CTCCAGT..CTCT..TC........GGGGG.CGTGGGTTCGAATCCCAC.CGCTGCCA...tRNA-Leu-AAG-2-3
GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTAAGG.CTCCAGT..CTCT..TC........GGGGG.CGTGGGTTCGAATCCCAC.CGCTGCCA...tRNA-Leu-AAG-2-4
GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTAAGG.CTCCAGT..CTCT..TC........GGGGG.CGTGGGTTCAAATCCCAC.CGCTGCCA...tRNA-Leu-AAG-3-1

GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTTAGG.CTCCAGT..CTCT..TC........GGAGG.CGTGGGTTCGAATCCCAC.CGCTGCCA...tRNA-Leu-TAG-1-1
GGTAGTGTGGCCGAGC.GGTct.A.AGGCGCTGGATTTAGG.CTCCAGT..CTCT..TC........GGGGG.CGTGGGTTCGAATCCCAC.CACTGCCA...tRNA-Leu-TAG-2-1
GGTAGCGTGGCCGAGT.GGTct.A.AGGCGCTGGATTTAGG.CTCCAGT..CATT..TC........GATGG.CGTGGGTTCGAATCCCAC.CGCTGCCA...tRNA-Leu-TAG-3-1
GGTAGCGTGGCCGAGC.GGTct.A.AGGCGCTGGATTTAGG.CTCCAGT..CTCT..TC........GGAGG.CGTGGGTTCGAATCCCAC.CGCTGCCA... DL9990_TAG_HUMAN_CY
```

**Leu2hn**

TGTCARAAGTGGGATTCGAACCCACGCCTCCAGWGGAGGWTGCGATTTGAACGCAGCGCCTTAAGACYGCTCGGCCATCC

tRNALeuCAG

```
>>>>>>>...>>>.............<<<..>>>>>.........<<<<<....>>>>.....<<<<...>>>>>.......<<<<<<<<<<<<<....
GTCAGGAT.GGCCGAGCGGTct.AAGGCG.CTGCG.TTCAGGT.CGCAGT...CTCC..CCTGGAGG.CGTGGGTTCGAATCCCACTCCTGACA...tRNA-Leu-CAG-1-7
GTCAGGAT.GGCCGAGCGGTct.AAGGCG.CTGCG.TTCAGGT.CGCAGT...CTCC..CCTGGAGG.CGTGGGTTCGAATCCCACTCCTGACA...tRNA-Leu-CAG-1-6
GTCAGGAT.GGCCGAGCGGTct.AAGGCG.CTGCG.TTCAGGT.CGCAGT...CTCC..CCTGGAGG.CGTGGGTTCGAATCCCACTCCTGACA...tRNA-Leu-CAG-1-5
GTCAGGAT.GGCCGAGCGGTct.AAGGCG.CTGCG.TTCAGGT.CGCAGT...CTCC..CCTGGAGG.CGTGGGTTCGAATCCCACTCCTGACA...tRNA-Leu-CAG-1-4
GTCAGGAT.GGCCGAGCGGTct.AAGGCG.CTGCG.TTCAGGT.CGCAGT...CTCC..CCTGGAGG.CGTGGGTTCGAATCCCACTCCTGACA...tRNA-Leu-CAG-1-3
GTCAGGAT.GGCCGAGCGGTct.AAGGCG.CTGCG.TTCAGGT.CGCAGT...CTCC..CCTGGAGG.CGTGGGTTCGAATCCCACTCCTGACA...tRNA-Leu-CAG-1-2
GTCAGGAT.GGCCGAGCGGTct.AAGGCG.CTGCG.TTCAGGT.CGCAGT...CTCC..CCTGGAGG.CGTGGGTTCGAATCCCACTCCTGACA...tRNA-Leu-CAG-1-1
GTCAGGAT.GGCCGAGCGGTct.AAGGCG.CTGCG.TTCAGGT.CGCAGT...CTCC..CCTGGAGG.CGTGGGTTCGAATCCCACTTCTGACA...tRNA-Leu-CAG-2-2
GTCAGGAT.GGCCGAGCGGTct.AAGGCG.CTGCG.TTCAGGT.CGCAGT...CTCC..CCTGGAGG.CGTGGGTTCGAATCCCACTTCTGACA... tRNA-Leu-CAG-2-1
GTCAGGGT.GGCTGAGCAGTct.GAGGGG.CTGCG.TTCAGGT.CGCAGT...CTGC..CCTGGAGG.CGTGGGTTCGAATCCCACTCCTGAAA... tRNA-Leu-CAG-chr9-7
```

**Leu3h**

TGTCAGAAGTGGGATTCGAACCCACGCCTCCATTGGAGACCAGAACTTGAGTCTGGCGCCTTAGACCACTCGGCCATCC

tRNALeuCAA

```
>>>>>>>...>>>.............<<<..>>>>>.........<<<<<....>>>>>..........<<<<<...>>>>>...........<<<<<....<<<<<<<....
GTCAGGAT.GGCCGAGTGGTct.AAGGCG.CCAGA.CTCAAGT.TCTGGT...CTCC..GTAT......GGAGG.CGTGGGTTCGAAT....CCCAC....TTCTGACA...tRNA-Leu-CAA-2-1
GTCAGGAT.GGCCGAGTGGTct.AAGGCG.CCAGA.CTCAAGT.TCTGGT...CTCC..GCAT......GGAGG.CGTGGGTTCGAAT....CCCAC....TTCTGACA...tRNA-Leu-CAA-3-1
GTCAGGAT.GGCCGAGTGGTct.AAGGCG.CCAGA.CTCAAGT.TCTGGT...CTCC..AAT.......GGAGG.CGTGGGTTCGAAT....CCCAC....TTCTGACA...tRNA-Leu-CAA-1-2
GTCAGGAT.GGCCGAGTGGTct.AAGGCG.CCAGA.CTCAAGT.TCTGGT...CTCC..AAT.......GGAGG.CGTGGGTTCGAAT....CCCAC....TTCTGACA...tRNA-Leu-CAA-1-1
GTCAGGAT.GGCCGAGTGGTct.AAGGCG.CCAGA.CTCAAGT.TCTGGT...CTCCGG........ATGGAGG.CGTGGGTTCGAAT....CCCAC....TTCTGACA...tRNA-Leu-CAA-4-1
```

**Leu4h**

TACCRGGAGTGGGGYTCGAACCCACGCGGACAYCSGTCCATTGGATCTTAAGTCCAACGCCTTAACCACTCRGCCATCC

tRNALeuTAA-1

```
>>>>>>>...>>>.............<<<..>>>>>.........<<<<<....>>>>.............<<<<...>>>>>.......<<<<<<<<<<<<<....
ACCAGGAT.GGCCGAGTGGTt..AAGGCG.TTGGA.CTTAAGA.TCCAAT...GGAC..ATAT......GTCCG.CGTGGGTTCGAACCCCACTCCTGGTA...tRNA-Leu-TAA-1-1
ACCAGAAT.GGCCGAGTGGTt..AAGGCG.TTGGA.CTTAAGA.TCCAAT...GGAT..TCAT......ATCCG.CGTGGGTTCGAACCCCACTTCTGGTA...tRNA-Leu-TAA-3-1
ACCGGGAT.GGCCGAGTGGTt..AAGGCG.TTGGA.CTTAAGA.TCCAAT...GGGCTG........GTGCCCG.CGTGGGTTCGAACCCCACTCTCGGTA...tRNA-Leu-TAA-2-1
ACCGGGAT.GGCTGAGTGGTt..AAGGCG.TTGGA.CTTAAGA.TCCAAT...GGACAG........GTGTCCG.CGTGGGTTCGAGCCCCACTCCCGGTA...tRNA-Leu-TAA-4-1
ACCAGGAT.GGCCAAGTAGTta.AAGGCA.CTGGA.CTTAAGA.GCCAAT...GGAC..ATAT......GTCTG.TGTGGGTTTGAACCCCACTCCTGGTG...tRNA-Leu-TAA-11-12
ACTCATTT.GGCTGAGTGGTt..AAGGCA.TTGGA.CTTAAGA.TCCAAT.G.GAGT..AGT.......GGCTG.TGTGGGTTTAAACCCCACTACTGGTA...tRNA-Leu-TAA-5-1
```

**Leu6h**

TGTTAAKRAGAGGAGTTGAACCTCTGATTATAAARTTTTAAGTYTTATGCAATTRCCRGGCTCTGCCATCTTAAC

tRNALeuTAA-2

```
>>>>>>>...>>>>............<<<<..>>>>>.........<<<<<.....>>>>>.......<<<<<<<<<<<<....
GTTAAGAT.GGCAGAGCCtGGTa..ATTGCA.TAAAA.CTTAAAA.TTTTATAAtCAGAGGTTCAACTCCTCTTCTTAACA... nmt-tRNA-Leu-TAA-1-1 Sc: 55.7
GTTAAGAT.GGCAGAGCCcGGCa..ATTGCA.TAAGA.CTTAAAA.CTTTATAAtCAGAGGTTCAACTCCTCTCATTAACA... nmt-tRNA-Leu-TAA-4-1 Sc: 30.0
```

**Lys1h**

CGCCCAACGTGGGGCTCGAACCCACGACCCTGAGATTAAGAGTCTCATGCTCTACCGACTGAGCTAGCCGGGC

tRNALysCTT

```
>>>>>>>..>>>>..........<<<<....>>>>>.......<<<<<......>>>>>.......<<<<<..<<<<<<<.
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGGGACTCTTAATCCCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG...tRNA-Lys-CTT-1-2
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGGGACTCTTAATCCCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG...tRNA-Lys-CTT-1-1
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGAGACTCTTAATCTCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG...tRNA-Lys-CTT-2-5
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGAGACTCTTAATCTCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG...tRNA-Lys-CTT-2-4
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGAGACTCTTAATCTCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG...tRNA-Lys-CTT-2-3
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGAGACTCTTAATCTCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG...tRNA-Lys-CTT-2-2
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGAGACTCTTAATCTCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG...tRNA-Lys-CTT-2-1
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGAGACCCTTAATCTCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG...tRNA-Lys-CTT-3-1
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGGGACTCTTAATCTCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG...tRNA-Lys-CTT-4-1
GCCCGGCTAGCTCAGTCGAT.A.GAGCA...TGAGACTCTTAATCTCAGGGtC.GTGGGTTCGAGCCGCAC..GTTGGGCG...tRNA-Lys-CTT-5-1
GCCCAGCTAGCTCAGTCGGT.A.GAGCA...TGAGACTCTTAATCTCAGGGtC.ATGGGTTTGAGCCCCAC..GTTTGGTG...tRNA-Lys-CTT-7-1
GACGAGCTAGCTCAGTCGGT.A.GAGCA...TGGGACTCTTAATCCCAGGGtC.GTGGGTTTGAGCCCCAT..GTTGGGCA...tRNA-Lys-CTT-6-1
GCCTGGCTAGCTCAGTCGGC.A.AAGCA...TGAGACTCTTAATCTCAGGGtC.GTGGGCTCGAGCTCCAT..GTTGGGCG...tRNA-Lys-CTT-8-1
CTGCAGCTAGCTCAGTCGGT.A.GAGCA...TGAGACTCTTAATCTCAGGGtC.ATGGGTTCGTGCCCCAT..GTTG..GG...tRNA-Lys-CTT-11-1
GCCCGACTACCTCAGTCGGT.G.GAGCA...TGGGACTCTTCATCCCAGGGtT.GTGGGTTCGAGCCCCAC..ATTGGGCA...tRNA-Lys-CTT-9-1
GCCCAGCTAGCTCAGTCGGT.A.GAGCA...TAAGACTCTTAATCTCAGGGtT.GTGGATTCGTGCCCCAT..GCTGGGTG...tRNA-Lys-CTT-10-1
```

```
CCAGCA.TGTCTCAGTCGGT.A.TAGTG...TGAGACTCTTAATCTCAGGGtC.GTGGGTTCAAGCCCCAC..A..TTGGG...tRNA-Lys-CTT-12-1
GCCCAGCTAGCTCAGCCGGT.A.GAGCA...CAAGACTCTTAATCTCAGGGtC.GTGGGTTTGAGCCCTGT..GTTGAGCA...tRNA-Lys-CTT-14-1
..CCGAATAGCTTAGTTGAT.G.AAGCG...TGAGACTCTTAATCTCAGGGtA.GTGGGTTCAAGCCCCAC..ATTGG..A...tRNA-Lys-CTT-15-1
GTCTAGCTAGATCAGTTGGT.A.GAGCA...TAAGACTCTTAATCTCAGGGtC.ATGGGTTTGAGCCCTAC..GTTGGGCG...tRNA-Lys-CTT-13-1
GCCTGGCTACCTCAGTTGGT.A.GAGCA...TGGGACTCTTAATCCCAGAGtCaGTGGGTTCAAGCCTCAC..ATTGAGTG...tRNA-Lys-CTT-16-1
CCCCGGCTGGCTCAGTCAGT.A.GATCA...TGAGACTCTTAATCTCAGGGtC.GTGGGTTCACGCCCCAC..ACTGGGCG...tRNA-Lys-CTT-17-1chr16
GCCCGGCTAGCTCAGTCGGT.A.GAGCA...TGAGACTCTTAATCTCAGGGtC.GTGGGTTCGAGCCCCAC..GTTGGGCG... DK9991_CTT_HUMAN_CY
```

**Lys2h**

CGCCYGAACAGGGACTTGAACCCTGGACCCTCAGATTAAAAGTCTGATGCTCTACCRACTGAGCTATCCRGGC

tRNALysTTT-1

```
>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<......>>>>>.......<<<<<..<<<<<<<.
GCCTGGATAGCTCAGTTGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCCTG..TTCAGGCA...tRNA-Lys-TTT-1-1
GCCTGGATAGCTCAGTTGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCCTG..TTCAGGCG...tRNA-Lys-TTT-2-1
GCCCGGATAGCTCAGTCGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCCTG..TTCGGGCG...tRNA-Lys-TTT-3-5
GCCCGGATAGCTCAGTCGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCCTG..TTCGGGCG...tRNA-Lys-TTT-3-4
GCCCGGATAGCTCAGTCGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCCTG..TTCGGGCG...tRNA-Lys-TTT-3-3
GCCCGGATAGCTCAGTCGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCCTG..TTCGGGCG...tRNA-Lys-TTT-3-2
GCCCGGATAGCTCAGTCGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCCTG..TTCGGGCG...tRNA-Lys-TTT-3-1
GCCTGGATAGCTCAGTCGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCCTG..TTCAGGCG...tRNA-Lys-TTT-4-1
GCCCGGATAGCTCAGTCGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CGGGGTTCAAGTCCCTG..TTCGGGCG...tRNA-Lys-TTT-5-1
GCCTGGGTAGCTCAGTCGGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCCTG..TCCAGGCG...tRNA-Lys-TTT-6-1
GCCTGGATAGCTCAGTTGGT.A.GAACATCAGACTTTTAATCTGACGGtG.CAGGGTTCAAGTCCCTG..TTCAGGCG...tRNA-Lys-TTT-7-1
GCCCGGAGAGCTCAGTGGGtA..GAGCATCAGACTTTTAATCTGAGGGtC.CAGGGTTCAAGTCCTCG..TTCGGGCA...tRNA-Lys-TTT-8-1
GCCAGGATAGTTCAGGTGGT.A.GAGCATCAGACTTTTAACCTGAGGGtT.CAGGGTTCAAGTCTCTG..TTTGGGCG...tRNA-Lys-TTT-10-1
GCCTGGGTAGCTCAGTCGGT.A.GAGCTTCAGACTTTTAGCCTGAGGAtT.CAGGGTTCAATCCCTTG..CTGGGGCG...tRNA-Lys-TTT-13-1
GCCTGGATAGCTCCTTCGGT.A.GAGCATCAGACTTTTAATGTGAGGGtC.CAGGGTTCAAGTTCCTG..TTTGGGCG...tRNA-Lys-TTT-chr3-2
```

**Lys3h**...

CACCCAAAAAGGGACATGAACCTTGGACCCTCAGATTAAAAGTCTGATGCTCTACTGACTGAGCTATCTGGGT

tRNALysTTT-2 ...

```
>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<.......>.>>.......<<.<...<<<<<<<.
ACCCAGATAGCTCAGTCAGT.A.GAGCATCAGACTTTTAATCTGAGGGtC.CAAGGTTCATGTCCCTT..TTTGGGTG...tRNA-Lys-TTT-11-1
```

**Lys4h**

CGCCTGGACAGGGACTTGAACCCTAGACCCTCAGATTAAAAGTCTGATGTTCTACCTACTGAGCTACCCAGGT

tRNALysTTT-3

```
.>>>>>>..>.>>..........<<.<.>>>>>.......<<<<<......>>>>>.......<<<<<..<<<<<<..
ACCTGGGTAGCTCAGTAGGT.A.GAACATCAGACTTTTAATCTGAGGGtC.TAGGGTTCAAGTCCCTG..TCCAGGCG...tRNA-Lys-TTT-9-1
ACCTGGGTAGCTTAGTTGGT.A.GAGCATTGGACTTTTAATTTGAGGGcC.CAGGTTTCAAGTCCCTG..TTTGGGTG...tRNA-Lys-TTT-12-1
```

**Met1h**

TAGCAGAGGATGGTTTCGATCCATCGACCTCTGGGTTATGGGCCCAGCACGCTTCCGCTGCGCCACTCTGCT

tRNAMet-i

```
>>>>>>>..>>>>..........<<<<.>...>>>>.......<<<<<.....>>>>>.......<<<<<<<<<<<<....
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-8
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-7
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-6
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-5
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-4
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-3
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-2
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-1
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCTAAACCATCCTCTGCTA...tRNA-iMet-CAT-2-1
AGCAGAGTGGTGCAGT.GGA.A.GCATACctaTGGGCCCATAACCCAGAGGtTGATGGATGGAAACCATCCTCTGCTA...tRNA-iMet-CAT-3-1
```

**Met2h**

TGCCCYCTCTGAGGCTCGAACTCAGGACCTTCAGATTATGAGACTGACGCGCTGCCHRCTGCGCTAAGRRGGC

tRNAMet-e

```
>>>>>>>..>>>>..........<<<<.>>>>.........<<<<.....>>>>>.......<<<<<<<<<<<<....
GCCTCGTTAGCGCAGTAGGT.A.GCGCGTCAGTCTCATAATCTGAAGGtCGTGAGTTCGATCCTCACACGGGGCA...tRNA-Met-CAT-1-1
GCCCTCTTAGCGCAG.TGGGcA.GCGCGTCAGTCTCATAATCTGAAGGtCCTGAGTTCGAGCCTCAGAGAGGGCA...tRNA-Met-CAT-2-1
GCCCTCTTAGCGCAG.CGGGcA.GCGCGTCAGTCTCATAATCTGAAGGtCCTGAGTTCGAGCCTCAGAGAGGGCA...tRNA-Met-CAT-4-3
GCCCTCTTAGCGCAG.CGGGcA.GCGCGTCAGTCTCATAATCTGAAGGtCCTGAGTTCGAGCCTCAGAGAGGGCA...tRNA-Met-CAT-4-2
GCCCTCTTAGCGCAG.CGGGcA.GCGCGTCAGTCTCATAATCTGAAGGtCCTGAGTTCGAGCCTCAGAGAGGGCA...tRNA-Met-CAT-4-1
GCCTCCTTAGCGCAGTAGGC.A.GCGCGTCAGTCTCATAATCTGAAGGtCCTGAGTTCGAACCTCAGAGGGGGCA...tRNA-Met-CAT-3-2
GCCTCCTTAGCGCAGTAGGC.A.GCGCGTCAGTCTCATAATCTGAAGGtCCTGAGTTCGAACCTCAGAGGGGGCA...tRNA-Met-CAT-3-1
GCCCTCTTAGCGCAGCTGGC.A.GCGCGTCAGTCTCATAATCTGAAGGtCCTGAGTTCAAGCCTCAGAGAGGGCA...tRNA-Met-CAT-5-1
GCCTCGTTAGCGCAGTAGGC.A.GCGCGTCAGTCTCATAATCTGAAGGtCGTGAGTTCGAGCCTCACACGGGGCA...tRNA-Met-CAT-6-1
GCCCTCTTAGTGCAGCTGGC.A.GCGCGTCAGTTTCATAATCTGAAAGtCCTGAGTTCAAGCCTCAGAGAGGGCA...tRNA-Met-CAT-7-1
```

**Phe1h**

TGCYGAAACCCGGGATYGAACCAGGGACCTTTAGATCTTCAGTCTAACGCTCTCCCAACTGAGCTATTTCGGC

tRNAPheGAA

```
>>>>>>>..>>>>..........<<<<.>>>>>.......<<<.<<......>>>>>.......<<<<<..<<<<<<<....
GCCGAAATAGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.AAAGGt.CCCTGGTTCGATCCCGGG..TTTCGGCA...tRNA-Phe-GAA-1-6
GCCGAAATAGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.AAAGGt.CCCTGGTTCGATCCCGGG..TTTCGGCA...tRNA-Phe-GAA-1-5
GCCGAAATAGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.AAAGGt.CCCTGGTTCGATCCCGGG..TTTCGGCA...tRNA-Phe-GAA-1-4
GCCGAAATAGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.AAAGGt.CCCTGGTTCGATCCCGGG..TTTCGGCA...tRNA-Phe-GAA-1-3
GCCGAAATAGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.AAAGGt.CCCTGGTTCGATCCCGGG..TTTCGGCA...tRNA-Phe-GAA-1-2
GCCGAAATAGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.AAAGGt.CCCTGGTTCGATCCCGGG..TTTCGGCA...tRNA-Phe-GAA-1-1
GCCGAGATAGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.AAAGGt.CCCTGGTTCAATCCCGGG..TTTCGGCA...tRNA-Phe-GAA-3-1
GCCGAAATAGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.AAAGGt.CCCTGGTTCAATCCCGGG..TTTCGGCA...tRNA-Phe-GAA-2-1
GCCGAAATAGCTCAGTTGGG.A.GAGCGTTAGACCGAAGATCTtAAAGGt.CCCTGGTTCAATCCCGGG..TTTCGGCA...tRNA-Phe-GAA-4-1
GCCAAAATTGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.AAAGGt.CCCTGGTTCGATCCCGGG..TTT...CA...tRNA-Phe-GAA-5-1
GCTGAAATAGCTCAGTTGGG.A.GAGCGTTAGACTGAAGATCT.TAAAGttCCCTGGTTCAACCCTGGG..TTTCAGCC...tRNA-Phe-GAA-6-1
GCCGAAATAGCTCAATTGGG.A.GAGTGTTAGACTGAAGTTCT.GCAGGt.CTCTGGTTCAATTCCGGG..TTTCGACA...tRNA-Phe-GAA-7-1
GCCAAAATAGCTCAGCTGGG.A.GAGCATTAGACTGAAGATCT.AAAGGt.CTCTGGTTTGATCCTGGG..TTT...CA...tRNA-Phe-GAA-8-1
```

**Pro1h**

GGGCTCGTCCGGGATTTGAACCCGGGACCTCTCGCACCCDAAGCGAGAATCATACCCCTAGACCAACGAGCC

tRNAProhGG

```
>>>>>>>..>>>.............<<<.>.>>>>.......<<<<<.....>>>>>.......<<<<<.<<<<<<<.
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGATGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-AGG-1-1
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-AGG-2-8
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-AGG-2-7
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-AGG-2-6
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-AGG-2-5
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-AGG-2-4
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-AGG-2-3
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-AGG-2-2
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-AGG-2-1
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTAGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...DP9990_AGG_HUMAN_CY

GGCAGTGTGGCATAGTGGTtAgaa.ATGTGC..GCTCTGGGGCTGC..TGAtCCCAGGCTCAAACCCTGG.CGCTGTCA...tRNA-Pro-GGG-1-1

GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTCGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-CGG-1-3
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTCGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-CGG-1-2
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTCGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-CGG-1-1
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTCGGGTGTGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-CGG-2-1

GGCTCGTTGGTCTAGTGGT.A....TGATTC.TCGCTTTGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-TGG-1-1
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGGTTTGGGTCCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-TGG-2-1
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTTGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-TGG-3-5
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTTGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-TGG-3-4
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTTGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-TGG-3-3
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTTGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-TGG-3-2
GGCTCGTTGGTCTAGGGGT.A....TGATTC.TCGCTTTGGGTGCGAGAGGtCCCGGGTTCAAATCCCGG.ACGAGCCC...tRNA-Pro-TGG-3-1
```

**Sec1h**

TGACCACAAAGGGACTCAAACCCTCAATCTTCTGATCTGAAGTCAGACACCTTATCCATTAAGGCCACACGGTC

tRNASec

```
>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<...............>>>>>.......<<<<<<<<<<<<.
GACCGTGTGGCCTtAATGGAtA.AGGTGTCTGACTTCAGATCAGAAG...........AtTGAGGGTTTGAGTCCCTTTGTGGTCA...tRNA-SeC-TCA-3-1
```

**Ser1h**

GCTGTGAGCAGGATTTGAACCTGCGCGGGGARACCCCATTGGATTTCGAGTCCAACGCCTTAACCACTCGGCCATCACAG

tRNASerCGA

```
>>>>>>>..>>>.............<<<.>>>>>.......<<<<<..>>>>......<<<<...>>>>>.......<<<<<...<<<<<<<....
GCTGTGATGGCCGAGT..GGTtA.AGGCGTTGGACTCGAAATCCAAT.GGGG...TCTCCCCG.CGCAGGTTCGAATCCTGC...TCACAGCG...tRNA-Ser-CGA-1-1
GCTGTGATGGCCGAGT..GGTtA.AGGCGTTGGACTCGAAATCCAAT.GGGG...TCTCCCCG.CGCAGGTTCAAATCCTGC...TCACAGCG...tRNA-Ser-CGA-2-1
GTCACGGTGGCCGAGT..GGTtA.AGGCGTTGGACTCGAAATCCAAT.GGGG...TTTCCCCG.CACAGGTTCGAATCCTGT...TCGTGACG...tRNA-Ser-CGA-4-1
GCTGTGATGGCCGAGT..GGTtA.AGGTGTTGGACTCGAAATCCAATGGGGG...TT.CCCCG.CGCAGGTTCAAATCCTGC...TCACAGCG...tRNA-Ser-CGA-3-1
```

**Ser2h**

GTAGYCGGCAGGATTCGAACCTGCGCGGGGARACCCCAATGGATTTCSAGTCCATCGCCTTAACCACTCGGCCACGACTA

tRNASerwGA

```
>>>>>>>..>>>...........<<<.>>>>........<<<<..>>>>...<<<<...>>>>>.......<<<<<<<<<<<<....
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTTTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-AGA-1-1
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-AGA-2-6
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-AGA-2-5
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-AGA-2-4
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-AGA-2-3
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-AGA-2-2
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-AGA-2-1
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTTTCCCCA.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-AGA-3-1
GTAGTCGTGGCCGAGTGGTtA.AGGTGATGGACTAGAAACCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-AGA-4-1
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTTTCCCCA.CGCAGGTTCGAATCCTGCCGACTACG...DS9991_AGA_HUMAN_CY
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTAGAAATCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...DS9992_AGA_HUMAN_CY

GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTTGAAATCCATT.GGGGTTTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-TGA-2-1
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTTGAAATCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...tRNA-Ser-TGA-3-1
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTTGAAATCCATT.GGGGTTTCCCCG.CGCAGGTTCGAATCCTGTCGGCTACG...tRNA-Ser-TGA-4-1
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTTGAAATCCATT.GGGGTTTCCCCG.CGCAGGTTCGAATCCTGCCGACTACG...DS9990_TGA_HUMAN_CY
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTTGAAATCCATT.GGGGTTTCCCCG.CGCAGGTTCGAATCCTGTCGGCTACG...DS9993_TGA_HUMAN_CY
GTAGTCGTGGCCGAGTGGTtA.AGGCGATGGACTTGAAATCCATT.GGGGTCTCCCCG.CGCAGGTTCGAATCCTGCCGACTACGcca RS9991_TGA_HUMAN_CY
```

**Ser3h**

GACGAGGRTGGGATTCGAACCCACGYGTGCAGAGCACAATGGATTAGCAGTCCATCGCCTTAACCACTCGGCCACCTCGT

tRNASerGCT

```
>>>>>>>..>>>...........<<<.>>>>>.......<<<<<..>>>>...<<<<...>>>>>.......<<<<<<<<<<<<....
GACGAGGTGGCCGAGTGGTtA.AGGCGATGGACTGCTAATCCATT.GTGCTCTGCACG.CGTGGGTTCGAATCCCACCCTCGTCG...tRNA-Ser-GCT-1-1
GACGAGGTGGCCGAGTGGTtA.AGGCGATGGACTGCTAATCCATT.GTGCTCTGCACG.CGTGGGTTCGAATCCCACCTTCGTCG...tRNA-Ser-GCT-2-1
GACGAGGTGGCCGAGTGGTtA.AGGCGATGGACTGCTAATCCATT.GTGCTTTGCACG.CGTGGGTTCGAATCCCATCCTCGTCG...tRNA-Ser-GCT-3-1
GACGAGGTGGCCGAGTGGTtA.AGGCGATGGACTGCTAATCCATT.GTGCTCTGCACG.CGTGGGTTCGAATCCCATCCTCGTCG...tRNA-Ser-GCT-4-3
GACGAGGTGGCCGAGTGGTtA.AGGCGATGGACTGCTAATCCATT.GTGCTCTGCACG.CGTGGGTTCGAATCCCATCCTCGTCG...tRNA-Ser-GCT-4-2
GACGAGGTGGCCGAGTGGTtA.AGGCGATGGACTGCTAATCCATT.GTGCTCTGCACG.CGTGGGTTCGAATCCCATCCTCGTCG...tRNA-Ser-GCT-4-1
GACGAGGTGGCCGAGTGGTtA.AGGCGATGGACTGCTAATCCATT.GTGCTCTGCACA.CGTGGGTTCGAATCCCATCCTCGTCG...tRNA-Ser-GCT-5-1
GAGG.CCTGGCCGAGTGGTtA.AGGCGATGGACTGCTAATCCATT.GTGCTCTGCACG.CGTGGGTTCGAATCCCAT...CCTCG...tRNA-Ser-GCT-6-1
```

**Ser4h**

AGGGGGCACCTGGATTTGAACCAGGGACCTCTTGATCTCTAGTCAAATTCTCTACCCCTGAGCCATACACCC

tRNASerAGA

```
>>>>>....>>>...........<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<..<<<<<....
GGGTGTATGGCTCAGGGGT.A.GAGAATTTGACTAGAGATCAAGAGGtCCCTGGTTCAAATCCAGGTGCCCCCT... tRNA-Ser-AGA-5-1 Sc: 53.2
```

tRNACysGCA (6 à 7 mismatches)

```
>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<.....>>>>>...........<<<<<...<<<<<<<.
GGGGGCATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-1-1
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-4-1
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-2-4
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-2-3
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-2-2
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-2-1
```

```
GGGGGTATAGCTCAG.GGGT..AGAGCACTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-3-1
GGGGGTATAGCTCAG.TGGGt.AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-5-1
GGGGGTGTAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-6-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-9-4
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-9-3
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-9-2
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-9-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCC...tRNA-Cys-GCA-10-1
GGGGGTATAGCTCAGTGGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-7-1
SGGGGGTATAGCTCAGGGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-8-1
SGGGGGTATAGCTTAGCGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-11-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCAG.TTCAAAT....CTGGG...TGCCCCCT...tRNA-Cys-GCA-13-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAAGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-14-1
GGGGGTATAGCTTAG.GGGT..AGAGCATTTGACTGCAGATCAAAAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCTT...tRNA-Cys-GCA-12-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCTCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-15-1
GGGGGTATAGTTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-18-1
GGGGGTATAGCTCAG.GGGT..AGAGCACTTGACTGCAGATCAAGAAGtCCTTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-16-1
...
```

**Thr1h**...

AGGCCCCRCTGGGAKTCGAACCCAGGATCTCCTGTTTACKAGACAGGCGC...TTTAACCARCTAAGCCAYRGNGCC

tRNAThrmGT

```
>>>>>>>..>>>>..........<<<<.>>>>>.......<<<.<<.....>>>>>...........<<<<<.<<<<<<<.
GGCGCCGTGGCTTAGTTGGTtA.AAGCGCCTGTCTAGTAAACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.CGGTGCCT...tRNA-Thr-AGT-1-3
GGCGCCGTGGCTTAGTTGGTtA.AAGCGCCTGTCTAGTAAACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.CGGTGCCT...tRNA-Thr-AGT-1-2
GGCGCCGTGGCTTAGTTGGTtA.AAGCGCCTGTCTAGTAAACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.CGGTGCCT...tRNA-Thr-AGT-1-1
GGCCCTGTGGCTTAGCTGGTcA.AAGCGCCTGTCTAGTAAACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.CGGGGCCT...tRNA-Thr-AGT-6-1
GGCTCCGTGGCTTAGCTGGTtA.AAGCGCCTGTCTAGTAAACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.CGGGGCCT...tRNA-Thr-AGT-2-2
GGCTCCGTGGCTTAGCTGGTtA.AAGCGCCTGTCTAGTAAACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.CGGGGCCT...tRNA-Thr-AGT-2-1
GGCGCCGTGGCTTAGCTGGTtA.AAGCGCCTGTCTAGTAAACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.CGGTGCCT...tRNA-Thr-AGT-5-1
GGCTCCGTAGCTTAGTTGGTtA.AAGCGCCTGTCTAGTAAACA.GGAGAtCCTGGG..TTCGACT...CCCAG.CGGGGCCT...tRNA-Thr-AGT-3-1
GGCTTCGTGGCTTAGCTGGTtA.AAGCGCCTGTCTAGTAAACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.CGAGGCCT...tRNA-Thr-AGT-4-1
AGCACCATGGCTTAGCTGGTtA.AAGCACCTGTCTAGTAAACA.GGAGAtCCTGAGt.TTCAATT....CCAA.TGGTGCCT...tRNA-Thr-AGT-7-1
...
GGCTCTATGGCTTAGTTGGTtA.AAGCGCCTGTCTCGTAAACA.GGAGAtCCTGGG..TTCGACT...CCCAG.TGGGGCCT...tRNA-Thr-CGT-1-1
GGCTCTGTGGCTTAGTTGGCtA.AAGCGCCTGTCTCGTAAACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.CGGGGCCT...tRNA-Thr-CGT-3-1
GGCAGAGTGGTGCAG.CGGA.A.GCGTGCTGGGCCCGTAACCC.AGAGGtCAATGG..ATCGAAG...CCATC.CTTGGCTA...tRNA-Thr-CGT-6-1
```

**Thr2h**

AGGCCCCAGCGAGATTYGAACTCGCGACCCCTGGTTTACAAGACCAGTGCTCTAACCMCTGAGCTATGGAGCC

tRNAThrTGT-1

```
>>>>>>>..>>>>..........<<<<.>>>>>........<<<.<<.....>>>>>...........<<<<<.<<<<<<<.
GGCTCTATGGCTTAGTTGGTtA.AAGCGCCTGTCTTGTA.AACA.GGAGAtCCTGGG..TTCGAAT...CCCAG.TAGAGCCT...tRNA-Thr-TGT-1-1
GGCTCCATAGCTCAG.TGGTtA.GAGCACTGGTCTTGTA.AACC.AGGGGtCGCGAG..TTCGATC...CTCGC.TGGGGCCT...tRNA-Thr-TGT-2-1
GGCTCCATAGCTCAG.GGGTtA.GAGCGCTGGTCTTGTA.AACC.AGGGGtCGCGAG..TTCAATT...CTCGC.TGGGGCCT...tRNA-Thr-TGT-3-1
GGCCCTATAGCTCAG.GGGTtA.GAGCACTGGTCTTGTA.AACC.AGGGGtCGCGAG..TTCAAAT...CTCGC.TGGGGCCT...tRNA-Thr-TGT-5-1
GGCTCCATAGCTCAG.GGGTtA.GAGCACTGGTCTTGTA.AACC.AGGGGtCGCGAG..TTCAAAT...CTCGC.TGGGGCCT...tRNA-Thr-TGT-4-1
GGCTCCATAGCTCAG.GGGTtA.GAGCACTGGTCTTGTA.AACC.AGG.GtCGCGAG..TTCAAAT...CTCGC.TGGGGCCT...tRNA-Thr-TGT-6-1
```

**Thr3h**

AGGCCCTGGTGAGATTTGAACTCACGACCCCTAGGTTTACGAGACCAGCGCTCCAACCGCTGAGCTACAGGGCC

tRNAThrCGT-1

```
>>>>>>>..>>>>..........<<<<.>>>>>........<<<.<<.....>>>>>...........<<<<<.<<<<<<<.
GGCCCTGTAGCTCAG.CGGTtG.GAGCGCTGGTCTCGTAAACCtAGGGGtCGTGAG..TTCAAAT...CTCAC.CAGGGCCT...tRNA-Thr-CGT-5-1
```

**Thr4h**

TGCTGCTCACAGGACTCGAACCTGCATTTCCATCCTTACAAGGGATGTGTAATAGCCAATTATACGAAAGCAGC

tRNAThrTGT-2

```
>>>>>>>..>>>>..........<<<<..>>>>>.......<<<<<.....>>>>>.......<<<<<<<<<<<.
GCTGCTTTCGTATAATT-GGCT-ATTACACATCCCTTGTAAGGATGGAAATGCAGGTTCGAGTCCTGTGAGCAGC...tRNA-Thr-TGT-1-1
```

**Thr5h**

AGGCACGGACGGGGTTCGAACCCGYGATCTTCGGTTTACGAGACCGACGCCTTACCACTTGGCCACCGCGCC

tRNAThrCGT-2

```
>>>>>>>..>>>.............<<<.>>>>>.......<<<.<<.....>>>>>...........<<<<<.<<<<<<<.
GGCGCGGTGGCCAAG.TGGT.A.AGGCGTCGGTCTCGTAAACC.GAAGAtCACGGG..TTCGAAC...CCCGT.CCGTGCCT...tRNA-Thr-CGT-2-1
GGCGCGGTGGCCAAG.TGGT.A.AGGCGTCGGTCTCGTAAACC.GAAGAtCGCGGG..TTCGAAC...CCCGT.CCGTGCCT...tRNA-Thr-CGT-4-1
...
```

**Thr6h**

TAGCAGAGGATGGTTTCGATCCATCGACCTCTGGGTTATGGGCCCAGCACGCTTCCGCTGCGCCACTCTGCT

ThrCGT

```
>>>>>>>..>>>>..........<<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<<<<<<<<.
GGCAGAGTGGTGCAGC.GGA.A.GCGTGCTGGGCCCGTAACCCAGAGGtCAATGGATCGAAGCCATCCTTGGCTA...tRNA-Thr-CGT-6-1
```

tRNAMet-i

```
>>>>>>>..>>>>..........<<<<.>...>>>.........<<<<.....>>>>>.......<<<<<<<<<<<<....
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-8
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-7
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-6
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-5
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-4
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-3
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-2
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...tRNA-iMet-CAT-1-1
GCCCTCTTAGTGCAGCTGGC.A.GCGCGT...CAGTTTCATAATCTGAAAGtCCTGAGTTCAAGCCTCAGAGAGGGCA...tRNA-Met-CAT-7-1
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCTAAACCATCCTCTGCTA...tRNA-iMet-CAT-2-1
AGCAGAGTGGTGCAGT.GGA.A.GCATACctaTGGGCCCATAACCCAGAGGtTGATGGATGGAAACCATCCTCTGCTA...tRNA-iMet-CAT-3-1
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...DM9990_CAT_HUMAN_CY
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCGAAACCATCCTCTGCTA...DX9990_CAT_HUMAN_CY
AGCAGAGTGGCGCAGC.GGA.A.GCGTGC...TGGGCCCATAACCCAGAGGtCGATGGATCTAAACCATCCTCTGCTA... DX9991_CAT_HUMAN_CY
```

## Tyr1h

AGGGGGCACCYGGATTTGAACCRGGGACCTCTTGATCTGCAGTCAAATGCTCTACCMCTGAGCTATACCCCC

### tRNACysGCA

```
>>>>.>>..>>>>..........<<<<.>>>>>.......<<<<<.....>>>>>...........<<<<<...<<.<<<<.
GGGGGCATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-1-1
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-4-1
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-2-4
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-2-3
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-2-2
GGGGGTATAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-2-1
GGGGGTATAGCTCAG.GGGT..AGAGCACTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-3-1
GGGGGTATAGCTCAG.TGGGt.AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-5-1
GGGGGTGTAGCTCAG.TGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-6-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-9-4
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-9-3
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-9-2
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-9-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCC...tRNA-Cys-GCA-10-1
GGGGGTATAGCTCAGGTGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-7-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-8-1
GGGGGTATAGCTTAG.CGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-11-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCAG.TTCAAAT....CTGGG...TGCCCCCT...tRNA-Cys-GCA-13-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAAGtCCCCGG.TTCAAAT....CCGGG...TGCCCCCT...tRNA-Cys-GCA-14-1
GGGGGTATAGCTTAG.GGGT..AGAGCATTTGACTGCAGATCAAAAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCTT...tRNA-Cys-GCA-12-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCTCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-15-1
GGGGGTATAGTTCAG.GGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-18-1
GGGGGTATAGCTCAG.GGGT..AGAGCACTTGACTGCAGATCAAGAAGtCCTTGG.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-16-1
GGGGGTATAGCTCAG.GGGT..AGAGCATTTGACTGCAAATCAAGAGGtCCCTGA.TTCAAAT....CCAGG...TGCCCCCT...tRNA-Cys-GCA-19-1
GGGGGTATAGCTCAC.AGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CTGGG...TGCCCCCT...tRNA-Cys-GCA-21-1
GGGGGTATAGCTCAC.AGGT..AGAGCATTTGACTGCAGATCAAGAGGtCCCCGG.TTCAAAT....CCGGT...TACTCCCT...tRNA-Cys-GCA-23-1
```

### tRNACysTGT...

```
>>>>>.>..>>>>..........<<<<..>>>>.........<<<<.....>>>>>...........<<<<<...<.<<<<<.
GGGGGTATAGCTCAG..GGCt.AGAGCTtTTTGACTGTAGAGCAAGAGGtCCCTGGTTCAAAT....CCAGG...TTCTCCCT...tRNA-Tyr-GTA-10-1 Sc: 49.8
```

## Tyr2h

TCCTTCGAGCCGGASTCGAACCAGCGACCTAAGGATCTACAGTCCTCCGCTCTACCARCTGAGCTATCGAAGG

### tRNATyrGTA

```
>>>>>>>..>>>>........<<<<..>>>>>.......<<<<<.....>>>>>.......<<<<<<<<<<<<....
CCTTCGATAGCTCAGTTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA...tRNA-Tyr-GTA-2-1
CCTTCGATAGCTCAGTTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGAATCCGGCTCGAAGGA...tRNA-Tyr-GTA-1-1
CCTTCGATAGCTCAGTTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGAATCCGGCTCGGAGGA...tRNA-Tyr-GTA-3-1
CCTTCGATAGCTCAGCTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA...tRNA-Tyr-GTA-5-5
CCTTCGATAGCTCAGCTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA...tRNA-Tyr-GTA-5-2
CCTTCGATAGCTCAGCTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGATTCCAGCTCGAAGGA...tRNA-Tyr-GTA-4-1
CCTTCGATAGCTCAGCTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA...tRNA-Tyr-GTA-5-1
CCTTCGATAGCTCAGCTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA...tRNA-Tyr-GTA-5-3
CCTTCGATAGCTCAGCTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGATTCCGGCTCGAAGGA...tRNA-Tyr-GTA-5-4
CCTTCGATAGCTCAGCTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGAATCCGGCTCGGAGGA...tRNA-Tyr-GTA-6-1
CCTTCGATAGCTCAGCTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCAATTCCGGCTCGAAGGA...tRNA-Tyr-GTA-7-1
CTTTCGATAGCTCAGTTGGTAGAGCG.GAGGACTGTAGATCCTTAGGtCGCTGGTTCGAATCCGGCTCGAAGGA... tRNA-Tyr-GTA-8-1
```

## Val1h

TGTTTCYGCCYGGTTTCGAACCRGGGACCTTTCGCGTGTKAGGCGAACGTGATAACCACTACACTACRGAAAC

### tRNAValmAC

```
>>>>>>>..>>>..........<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<<<<<<<<....
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-AAC-1-5
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-AAC-1-4
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-AAC-1-3
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-AAC-1-2
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-AAC-1-1
GTTTCCGTAGTGTAGTGGTcATCACGTTCGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-AAC-2-1
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCCTGGATCAAAACCAGGCGGAAACA...tRNA-Val-AAC-3-1
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCGCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-AAC-4-1
GTTTCCGTAGTGTAGTGGTtATCACGTTTGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCAGAAACA...tRNA-Val-AAC-5-1
GTTTCCATAGTGTACTGGTtATCACATTCACCTAACACGCGAAAGGtCCTTGGTTTGAAACCAGGCAGAAACA...tRNA-Val-AAC-7-1
GTTTCCATAGCGTAGTAATtATCAAGTTTGCCTAACACGCGAAATGtCCCCGGTTCAAAACCGGGCGGAAACA...tRNA-Val-AAC-chr6-84
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...DV9990_AAC_HUMAN_CY
GTTTCCGTAGTGTAGTGGTcATCACGTTCGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...DV9992_AAC_HUMAN_CY
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCCTGGATCAAAACCAGGCGGAAACA...DV9993_AAC_HUMAN_CY
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTAACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...DV9996_AAC_HUMAN_CY

GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-CAC-1-6
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-CAC-1-5
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-CAC-1-4
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-CAC-1-3
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-CAC-1-2
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-CAC-1-1
GCTTCTGTAGTGTAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTCGAAACCGGGCAGAAGCA...tRNA-Val-CAC-2-1
GTTTCCGTAGTGTAGCGGTtATCACATTCGCCTCACACGCGAAAGGtCCCCGGTTCGATCCCGGGCGGAAACA...tRNA-Val-CAC-3-1
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTCGAAACTGGGCGGAAACA...tRNA-Val-CAC-4-1
GTTTCCGTAGTGTAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTCGAAACCGGGCGGAAACA...tRNA-Val-CAC-5-1
GTTTCCGTAGTGGAGTGGTtATCACGTTCGCCTCACACGCGAAAGGtCCCCGGTTTGAAACCAGGCGGAAACA...tRNA-Val-CAC-6-1
GTTTCCGTAGTGTAGTGGTtATTATGTTCGCCTCACACGCGAAAAGtCCCCGGTTCGAAATCAGGCGGGAACA...tRNA-Val-CAC-7-1
GTTTCTGTAGTATGGTGGTtATCACGTTAGTCTCACACGTGAAAGGtCCCTGGTTCGAAACCAGGTGGAAACA...tRNA-Val-CAC-9-1
GTTTCTGTGGTGTAGTGGTtATCATGTTCGCCTCACACGAGAAAAGtCCCTGATTCGAGACTGGGTGGGAACG...tRNA-Val-CAC-10-1
GTTTCTGTGGTGTAGTGGTtATTATGTTCGCTTCACATATGAAAGGtCTCTGGTTCGAGACTGCGTGGGAACA...tRNA-Val-CAC-11-2
GTTTCTGTGGTGTAGTGGTtATTATGTTCGCTTCACATATGAAAGGtCTCTGGTTCGAGACTGCGTGGGAACA...tRNA-Val-CAC-11-1
```

## Val3h

TGGAA*G*TGCTGGGGAT*C*GAACCCAGA*G*CCTC*A*T*G*AATGT*T*AAGCA*T*ACGCTCTACCACTGAGCTACA*C*CCCC

### tRNAAlaTGC

```
>>>>>>>..>>>>...........<<<<.>>>>>.......<<<<<......>.>>.........<<.<.<<<<<<<....
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGtCCCGGG.TTCGATC.CCCGGCACCTCCA...tRNA-Ala-TGC-1-1
GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGtCCCGGG.TTCGATC.CCCGGCATCTCCA...tRNA-Ala-TGC-2-1
GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGcCCCGGG.TTCGATC.CCCGGCATCTCCA...tRNA-Ala-TGC-3-2
GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGcCCCGGG.TTCGATC.CCCGGCATCTCCA...tRNA-Ala-TGC-3-1
GGGGATGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCACGTATGAGGcCCCGGG.TTCAATC.CCCGGCATCTCCA...tRNA-Ala-TGC-4-1
```

```
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGGcCTCGGG.TTCGATC.CCCGACACCTCCA...tRNA-Ala-TGC-5-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCACATGCTTTGCATGTGTGAGGcCCCGGG.TTCGATC.CCCGGCACCTCCA...tRNA-Ala-TGC-6-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCGCATGCTTTGCATGTATGAGG.CCTCGG.TTCGATC.CCCGACACCTCCA...tRNA-Ala-TGC-7-1
GGGGGTGTAGCTCAGT.GGT.A..GAGCGGATGCTTTGCATGTATGAGAcTTTGGG.TTGGATC.CCCAGCACCTCCA...tRNA-Ala-TGC-8-1
GTGGATGTAGTTTAGT.GGT.A..GAACGCGCGCTTTGCATGTATGAGGtCCCGGT.TTCGATC.CCTGGCGTTTCCA...tRNA-Ala-TGC-9-1
tRNAValAAC
>>>>>>>..>>>>.........<<<<.>>>>.........<<.<<...............>>>>>.......<<<<<<<<<<<<....
GGGGGTGTAGCTCAGTGGT.A.GAGCGTATGCTTAACATTCA.TGAG...........GcTCTGGGTTCGATCCCCAGCACTTCCA...tRNA-Val-AAC-6-1

Val4h
TGGTTCCACTGGGGCTCGAACCCAGGACCTTCTGCGTGTAAAGCAGACGTGATAACCACTACACTATGGAACC
tRNAValTAC
>>>>>>>..>>>.............<<<.>>>>>.......<<<.<<.....>>>>>.......<<<<<.<<<<<<<....
GGTTCCATAGTGTAG..TGGTtA..TCACGTCTGCTTTACACGCA.GAAGGtCCTGGGTTCGAGCCCCAG.TGGAACCA...tRNA-Val-TAC-1-2
GGTTCCATAGTGTAG..TGGTtA..TCACGTCTGCTTTACACGCA.GAAGGtCCTGGGTTCGAGCCCCAG.TGGAACCA...tRNA-Val-TAC-1-1
GGTTCCATAGTGTAG..CGGTtA..TCACGTCTGCTTTACACGCA.GAAGGtCCTGGGTTCGAGCCCCAG.TGGAACCA...tRNA-Val-TAC-2-1
GGTTCCATAGTGTAG..TGGTtA..TCACATCTGCTTTACACGCA.GAAGGtCCTGGGTTCAAGCCCCAG.TGGAACCA...tRNA-Val-TAC-3-1
GTTTCCGTGGTGTAG..TGGTtA..TCACATTCGCCTTACACGCG.AAAGGtCCTCGGGTCGAAACCGAG.CGGAAACA...tRNA-Val-TAC-4-1
GGTTCCATAGTGTAA..CAGTtA..TCATGTCTGCTTTACACACT.GAAGGtCCTGGGTTCAAACTCCAG.TGGAACCA...tRNA-Val-TAC-chr2-6
GTTAAGACGGCAGAGCCcGGCaA..TTGCGTAAAATTTACAACTT.TATGGgCAGAGGTTCAATTCCTCT.TCTTAACA...nmt-tRNA-Val-TAC-1-1
```

**r : A,G  y : C,T  m : A,C  s : C,G  w : A,T  h : A,C,T**

**Figure Supplémentaire 1 : Liste des gènes d'ARNt humains reconnus par les sondes utilisées pour le MIST.**

```
>ARNt^Asn_GTT (Asn-1h)
GTCTCTGTGGCGCAATCGGCTAGCGCGTTTGGCTGTTAACTAAAAGGTTGGTGGTTCGAACCCACCCAGAGGCG
>ARNt^Asp_GTC (Asp-1h)
TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
>ARNt^Lys_TTT-1 (Lys-2h)
GGCCGGATAGCTCAGTCGGTAGAGCATCAGACTTTTAATCTGAGGGtCCAGGGTTCAAGTCCCTGTTCGGCCG
>ARNT^Tyr_GTA (Tyr-2h)
CCTTCGATAGCTCAGCTGGTAGAGCGGAGGACTGTAGATCCTTAGGTCGCTGGTTCAATTCCGGCTCGAAGGA
>ARNt^Ala_AGC (Ala-1h)
GGGGGTGTAGCTCAGTGGTAGAGCGCGTGCTTAGCATGCACGAGGCCCCGGGTTCAATCCCCGGCACCTCCACCA
>ARNt^Ser_AGA (Ser-2h)
GTAGTCGTGGCCGAGTGGTTAAGGCGATGGACTAGAAATCCATTGGGGTTTCCCCGCGCAGGTTCGAATCCTGCCGACTACG
>ARNt^Thr_CGT-2(Thr-5h)
GGCGCGGTGGCCAAGTGGTAAGGCGTCGGTCTCGTAAACCGAAGAtCACGGGTTCGAACCCCGTCCGTGCCTCCA
>ARNt^Tyr_GCA(Tyr-1h)
GGGGGCATAGCTCAGTGGTAGAGCATTTGACTGCAGATCAAGAGGtCCCTGGTTCAAATCCAGGTGCCCCCTCCA
>ARNt^Arg_ACG (Arg-1h)
GGGCCAGTGGCGCAATGGAtAACGCGTCTGACTACGGATCAGAAGAtTCCAGGTTCGACTCCTGGCTGGCTCGCCA
>ARNt^Leu_TAG (Leu-1h)
GGTAGCGTGGCCGAGCGGTctAAGGCGCTGGATTTAGGCTCCAGTCTCTTCGGAGGCGTGGGTTCGAATCCCACCGCTGCCACCA
```

**Figure Supplémentaire 2 : Séquences des transcrits d'ARNt sélectionnés**

**Mm-GluCTC** : ARNt$^{Glu}_{CTC}$ souris (AATCCTAACCACTAGACCACCAGGGA)

**Pb-GluCTC** : ARNt$^{Glu}_{CTC}$ Pberhei (CCTTTCGGGTGAGAGCCGAATATCC)

```
berghei ARNtGluCTC                    TCCCACGTGGTCTAGTGGCTAGGATATTCGGCTGTCACCCGAAAGGCCCGGGTTCAATTCCCGGCGTGGGAA
Mus_musculus_chr3.trna622-GluCTC      TCCCTGATGGTATAGTGGTTAGGACTCGGTGGTCTCACCAGCGCTGCCCGGGTTCAATTCCTGGTTAGGGAA
Mus_musculus_chr12.trna892-GluCTC     TCCCTGGTGGTCTAGTGGTTAGGAGTCATTGCTCTCACCACCGCGTCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr7.trna969-GluCTC      TCCCTGGTGGTCTAGTGGTTAGGCTTTGGTGCTCTCACCTCCATGGCCCAGGTTTGATTCCTGGTCAGGGAA
Mus_musculus_chr17.trna719-GluCTC     TCCCTGGTGGTCTAGTGGTTAGGATTTGGCGCTCTCACCGCCGCGGCCTGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr10.trna90-GluCTC      TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr1.trna1004-GluCTC     TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr1.trna1007-GluCTC     TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr1.trna1010-GluCTC     TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr11.trna1912-GluCTC    TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr13.trna1013-GluCTC    TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr1.trna709-GluCTC      TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr3.trna303-GluCTC      TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr3.trna745-GluCTC      TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr3.trna286-GluTTC      TCCCTGGTGGTCTAGTGGCTAGGATTCGGCGCTTTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
Mus_musculus_chr3.trna754-GluTTC      TCCCTGGTGGTCTAGTGGCTAGGATTCGGCGCTTTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGGAA
```

**Mm-GluTTC** : ARNt$^{Glu}_{TTC}$ souris (GCCGCCTGGGTGAAAACCAGGAA)

**Pb-GluTTC** : ARNt$^{Glu}_{TTC}$ Pberghei (GTCGTTCGGGTGAAAGCCGAATATCC)

```
berghei ARNtGluTTC                    TCCCACGTAGTCTAGACGGTTAGGATATTCGGCTTTCACCCGAACGACCCGGGTTCGAGTCCCGGCGTGGGAA
Mus_musculus_chr1.trna1555-GluTTC     TCCCATATGGTCTAG-CGGTTAGGATTCCTGGTTTTCACCCAGGCGGCCCGGGTTCGACTCCCGGTATGGGAA
Mus_musculus_chr14.trna364-GluTTC     TCCCATATGGTCTAG-CGGTTAGGATTCCTGGTTTTCACCCAGGCGGCCCGGGTTCGACTCCCGGTATGGGAA
Mus_musculus_chr13.trna105-GluTTC     TCCCACATGGTCTAG-CGGTTAGGATTCCTGGTTTTCACCCAGGCGGCCCGGGTTCGACTCCCGGTGTGGGAA
Mus_musculus_chr14.trna352-GluTTC     TCCCACATGGTCTAG-CGGTTAGGATTCCTGGTTTTCACCCAGGCGGCCCGGGTTCGACTCCCGGTGTGGGAA
Mus_musculus_chr7.trna339-GluTTC      TCCCACATGGTCTAG-CGGTTAGGATTCCTGGTTTTCACCCAGGCGGCCCGGGTTCGACTCCCGGTGTGGGAA
Mus_musculus_chr9.trna961-GluTTC      TCCCACATGGTCTAG-CGGTTAGGATTCCTGGTTTTCACCCAGGCGGCCCGGGTTCGACTCCCGGTGTGGGAA
```

**Mm-Asp**: ARNt$^{Asp}$ souris (GCGGGGATACTCACCACTATACTAACGAGGA)

```
berghei ARNtAsp                       TCCGAGATAGTATAGTGGCAAGTATTTCCGCCTGTCACGCGGAAGACCCGGGTTCAATTCCCGGTCTCGGAG
Mus_musculus_chr2.trna211-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCTGCCTGTCACGCAGGACACCAGGGTTCGATTTCCTGACGGGGAG
Mus_musculus_chr5.trna719-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACACAGGAGACCGGGGTTCGATTCCCCGATGGGGAG
Mus_musculus_chr1.trna450-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr13.trna48-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr1.trna443-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr1.trna440-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr1.trna297-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr5.trna588-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr10.trna379-AspTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr11.trna29-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr5.trna586-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr1.trna446-AspGTC      TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr13.trna472-AspGTC     TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr13.trna474-AspGTC     TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr11.trna189-AspGTC     TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAG
Mus_musculus_chr10.trna383-AspGTC     TCCTCGTTAGTATAGTGGTTAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCAATTCCCCGACGGGGAG
```

## Mm-Asn : ARNt$^{Asn}$ souris (CAACCTTTCGGTTAACAGCCGAACGCGC)

```
Pberghei ARNtAsn                    GGTTCCGTAGCTCAGTTGGTTAGAGCGTGCGGCTGTTAACCGCAAGGTCGTTGGTTCGATCCCAGCCGGTACCG
Mus_musculus_chr3.trna350-AsnGTT    GTCTCTGTGGCGCAATCGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCAGGGACG
Mus_musculus_chr3.trna128-AsnGTT    GTCTCTGTGGCGCAATCGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCAGGGACG
Mus_musculus_chr1.trna451-AsnGTT    GTCTCTGTGGCGCAATCGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCAGGGACG
Mus_musculus_chr2.trna819-AsnGTT    GTCTCTGTGGCGCAATCGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCAGGGACG
Mus_musculus_chr3.trna134-AsnGTT    GTCTCTGTGGCGCAATCGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCAGGGACG
Mus_musculus_chr3.trna138-AsnGTT    GTCTCTGTGGCGCAATCGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCAGGGACG
Mus_musculus_chr3.trna343-AsnGTT    GTCTCTGTGGCGCAATCGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCAGGGACG
Mus_musculus_chr3.trna141-AsnGTT    GTCTCTGTGGCGCAATCGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCAGGGACG
Mus_musculus_chr5.trna472-AsnGTT    GTCTCTGTGGCGCAATCGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCAGGGACG
Mus_musculus_chr10.trna167-AsnGTT   GTCTCTGTGGCGCAATTGGTTAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCAAGCCCACCCAGGGACG
Mus_musculus_chr11.trna632-AsnGTT   GTCTCTGTGGCGCGATCGGTTAGTGCGTTTGGCTGTTAACCGAAAGGTTGGTGGTTCAAACCCACCCGGGGACG
Mus_musculus_chr15.trna437-AsnGTT   GTCTCCGTGGCGCAATCGGTCAGCGCGTTCGGCTGTTAACCGAAAGGTTGGTGGTTCGAGCCCACCCGGGGACG
```

## Mm-HisGTG : ARNt$^{His}{}_{GTT}$ souris souris (GCAGAGTACTAACCACTATACGATCACGGC)

## Pb-HisGTG : ARNt$^{His}{}_{GTT}$ Pberghei (GTGTGGAGTCCTAACCACTAGACGATTTGGA)

```
berghei ARNtHisGTG                  GTCCAAATCGTCTAGTGGTTAGGACTCCACACTGTGGATGTGGCAACGCAGGTTCGAATCCTGCTTTGGACA
Mus_musculus_chr3.trna293-HisGTG    GCCGTGATCGTATAGGGGTTAGTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA
Mus_musculus_chr2.trna1431-HisGTG   GCCGTGATCGTATAGTGGTTAGTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA
Mus_musculus_chr2.trna1432-HisGTG   GCCGTGATCGTATAGTGGTTAGTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA
Mus_musculus_chr2.trna587-HisGTG    GCCGTGATCGTATAGTGGTTAGTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA
Mus_musculus_chr3.trna291-HisGTG    GCCGTGATCGTATAGTGGTTAGTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA
Mus_musculus_chr3.trna295-HisGYG    GCCGTGATCGTATAGTGGTTAGTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA
Mus_musculus_chr3.trna747-HisGTG    GCCGTGATCGTATAGTGGTTAGTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA
Mus_musculus_chr3.trna751-HisGTG    GCCGTGATCGTATAGTGGTTAGTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA
Mus_musculus_chr4.trna1691-HisGTG   GCCGTGATCGTATAGTGGTTAGTACTCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCA
Mus_musculus_chr3.trna749-HisGTG    GCCGAGATCGTATAGTGGTTAGTACTCTGCATTGTGGCTGCAGCAACCTCGGTTCGAATCCGAGTCTCGGCA
```

## Mm-Phe : ARNt$^{Phe}$ souris (CAGTCTAACGCTCTCCCAACTGAGCTATTTC)

```
berghei ARNtPheGAA                  GCCGTGATAGCTCAGTTGGGAGAGCGTCAGACTGAAGATCTGAAGGTCCCTGGTTCGATCCCTGGTCACGGCA
Mus_musculus_chr5.trna587-PheG      GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAGATCTAAAGGTCCCTGGTTCGATCCCGGGTTTCGGCA
Mus_musculus_chr19.trna48-PheG      GCTGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAGATCTAAAGGTCCCTGGTTCGATCCCGGGTTTCAGCA
Mus_musculus_chr19.trna46-PheG      GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAGATCTAAAGGTCCCTGGTTCGATCCCGGGTTTCGGCA
Mus_musculus_chr14.trna204-Phe      GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAGATCTAAAGGTCCCTGGTTCGATCCCGGGTTTCGGCA
Mus_musculus_chr13.trna25-PheG      GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAGATCTAAAGGTCCCTGGTTCGATCCCGGGTTTCGGCA
Mus_musculus_chr10.trna427-Phe      GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAGATCTAAAGGTCCCTGGTTCGATCCCGGGTTTCGGCA
Mus_musculus_chr13.trna480-Phe      GCCGAAATAGCTCAGTTGGGAGAGCGTTAGACTGAAGATCTAAAGGTCCCTGGTTCAATCCCGGGTTTCGGCA
```

## Mm-Meti : ARNt$^{Met}$ initiateur souris (GCACGCTTCCGCTGCGCCACTCTGCT)

## Pb-Meti : ARNt$^{Met}$ initiateur Pberghei (CGCGTGGTTTCGATCCACGGTCCTGG)

```
berghei ARNtMeti                    AGCAGCGTAGCTCAGAGGAAGAGTGGGGGGCTCATAACCCCCAGGACCGTGGATCGAAACCACGCGCTGCTA
Mus_musculus_chr13.trna82-MetC      AGCAGAGTGGCGCAGCGGAAGCGTGCTGGGCCCATAACCCAGAGGTCGATAGATCGAAACCATCCTCTGCTA
Mus_musculus_chr15.trna876-Met      AGCAGAGTGGCGCAGCGGAAGCATGCTGGGCCCATAACCCAGAGGTCGATGGATCGAAACCATCCTCTGCTA
Mus_musculus_chr13.trna108-Met      AGCAGAGTGGCGCAGCGGAAGCGTGCTGGGCCCATAACCCAGAGGTCGATGGATCGAAACCATCCTCTGCTA
```

```
Mus_musculus_chr13.trna111-Met      AGCAGAGTGGCGCAGCGGAAGCGTGCTGGGCCCATAACCCAGAGGTCGATGGATCGAAACCATCCTCTGCTA
Mus_musculus_chr13.trna78-MetC      AGCAGAGTGGCGCAGCGGAAGCGTGCTGGGCCCATAACCCAGAGGTCGATGGATCGAAACCATCCTCTGCTA
Mus_musculus_chr13.trna90-MetC      AGCAGAGTGGCGCAGCGGAAGCGTGCTGGGCCCATAACCCAGAGGTCGATGGATCGAAACCATCCTCTGCTA
Mus_musculus_chr13.trna959-Met      AGCAGAGTGGCGCAGCGGAAGCGTGCTGGGCCCATAACCCAGAGGTCGATGGATCGAAACCATCCTCTGCTA
Mus_musculus_chr13.trna995-Met      AGCAGAGTGGCGCAGCGGAAGCGTGCTGGGCCCATAACCCAGAGGTCGATGGATCGAAACCATCCTCTGCTA
Mus_musculus_chr3.trna792-MetC      AGCAGAGTGGCGCAGCGGAAGCGTGCTGGGCCCATAACCCAGAGGTCGATGGATCGAAACCATCCTCTGCTA
```

**Mm-Mete :** ARNt$^{Met}$ élongateur souris (GGACCTTCAGATTATGAGACTGACGCGCT)

**Pb-Mete :** ARNt$^{Met}$ élongateur Pberghei (GCGCGCTAACCGACTGCGCCAATTACCC)

```
berghei ARNTMete                    GGGTAATTGGCGCAGTCGGTTAGCGCGCGGGTCTCATAATCCCGAGGTCGTGAGTTCGAGCCTCACATTACCCA
Mus_musculus_chr13.trna99-MetC      GCCTTCTTAGCGCAGTAGGC-AGCGCGTCAGTCTCATAATCTGAAGGTCCTGAGTTCGAGCCTCAGAGAAGGCA
Mus_musculus_chr13.trna1006-Me      GCCTTCTTAGCGCAGTAGGC-AGCGCGTCAGTCTCATAATCTGAAGGTCCTGAGTTCGAACCTCAGAGAGGGCA
Mus_musculus_chr13.trna1011-Me      GCCTCCTTAGCGCAGTAGGC-AGCGCGTCAGTCTCATAATCTGAAGGTCCTGAGTTCGAACCTCAGAGGGGGCA
Mus_musculus_chr13.trna61-MetC      GCCTCCTTAGCGCAGTAGGC-AGCGCGTCAGTCTCATAATCTGAAGGTCCTGAGTTCGAACCTCAGAGGGGGCA
Mus_musculus_chr15.trna913-Met      GCCTCGTTAGCGCAGTAGGT-AGCGCGTCAGTCTCATAATCTGAAGGTCGTGAGTTCGATCCTCACACGGGGCA
Mus_musculus_chr8.trna788-MetC      GCCTCGTTAGCGCAGTAGGT-AGCGCGTCAGTCTCATAATCTGAAGGTCGTGAGTTCGATCCTCACACGGGGCA
Mus_musculus_chr11.trna1132-Me      GCCTCCTTAGTGTAGTAGGC-ATTGCGTCAGTCTCATAATCTGAAGGTCCTGAGTTCAAGCCTCAGAGTGGGCA
Mus_musculus_chr19.trna371-Met      GCCTCCTTAGTGTAGTAGGC-ATTGCGTCAGTCTCATAATCTGAAGGTCCTGAGTTCAAGCCTCAGAGTGGGCA
Mus_musculus_chr15.trna355-Met      GCCTCCTTAGCGCAGTAGGC-AGCCCGTCAGTCTCATAATCTGAAGGTCCTGAGTTCAAACCTCAGAGGGGGCA
```

**Mm-Sec :** ARNt$^{Sec}$ souris (GCCTGCACCCCAGACCACTGAGGA)

**Pb-Sec :** ARNt$^{Sec}$ Pberghei (CTCTGTGCAGTGACCCTACACCAAATGA)

```
PbergheiARNtSec       GCATCGATGAGTTAGCATGGTTGCTAAAGATGACTTCAAATCATTTGGTGTAGGGTCACTGCACAGAGGTTCGATTCCTCCTTCGGTGC
Mus-musculusARNtSec   GCCCGGATGATCCTCAGTGGTCTGGGGTGCAGGCTTCAAACCTGTAGCTGTTTAGCGACAG---AGTGGTTCAATTCCACCTTTCGGGC
```

**ARN5s :** rRNA5s souris/Pberghei (gtgttcccgagcggtctcccacctc)

```
berghei5SrRNA   TGACTCGTTCATACTACAGTGGCCACACCAGATCCCATCAGAACTCTGAAGTTAAGCACTGTAAGGC-TTGGCTAGTACTGAGGTGGGAGACCGCTCGGGAACACTAAGTGATG-AGTCAT-
mus_musculus_5S  TCTACGGCCATACCACCCTGAACGCGCCCGATCTCGTGTGATGTCGGAAGCTAAGCAGGGTCGGGCCTTGGTTAGTACTTGGATGGGAGACCGCCTGGGAATACCGGGTGCTGTAGGCTTT
```

**Mm-7SL :** RNA 7SL souris (TGGTCCCCCGCTCCCGGGAGGTCACCATATTG)

**Pb-7SL :** RNA 7SL Pberghei (CTACAGCATGTTATCTTTCCAACATAACATGCTCAGTAACTTA)

```
berghei7SL        ------TCCGGCTGTATGCTCCTCGGGGTATGCAGCTGTAACCGTAAGG---GGAAAGCAAAGTAATTTGCG--CAGCCTTTAAGGGATGT--TACTC-ATGCCACGAGAAATTCAATAT
Mus-musculus7SL   GCCGGGCGCGGTGGCGCACGCCT-GTAGTCC-CAGCTAC--TCGGGAGGCTGAGACAGGAGGATCGCTTGAGTCCAGGAGTTCTGGGCTGTAGTGCGCTATGCCGATCGGGTGTCCGCAC

berghei7SL        ATAAGAGCTCTCGTCTATTGCAAGATA-TAAGTTACTGAGCATGTTATGTTGGAAAGATAACATGCTGTAGTAATAATGAATTGCGGAACGCCTTAGGCTGGAAACAGAGCAGGGTTAGC
Mus-musculus7SL   -TAAG----TTCGGC-AT--CAATATGGTGACCTCCCGGGAGCG--------GGGGACCACCAGGTTGCC----TAAGGAGGGGTGAACCGGCCCAGGTCGGAAACGGAGCAGG-TCAAA

berghei7SL        AATCGCTTGTATA-CGGTGCAAAGTTTGCG---ATGAGCGATTAGTATATCTCGAATCTCGTCAAGCATGGTAAGCACACCAATTGG
Mus-musculus7SL   ACTCCCGTGCTGATCAGTAGTGGGATCGCGCCTGTGAATAGCCACTGCA-CTCCAGCCTGGGCAA-CATAGCGAG-ACCCCGTCTCT
```

**Figure Supplémentaire 3 : Séquences des sondes utilisées dans les expériences de Northern.** Les sondes dirigées contre les ARNt de souris sont surlignées en jaune et les sondes correspondant aux ARNt de *Plasmodium berghei* en rose.

# CONCLUSION ET PERSPECTIVES

Comme tous les parasites, *Plasmodium* interagit avec son organisme hôte pour se développer de façon optimale. Ce parasite intracellulaire a développé plusieurs stratégies pour échanger toutes sortes de molécules avec son hôte, telles que des nutriments, dont l'hémoglobine (Revu dans Bombardier et Munson 2015), le glucose, les anions (Revu dans Landfear, 2011 et Kirk et Lehane, 2014), ou l'export de protéines (Revu dans Spillman, *et al.*, 2015). C'est dans ce contexte que le transport intracellulaire des ARNt, entre la cellule hôte et le parasite, a été décrit pour la première fois dans notre laboratoire (Bour, *et al.*, 2016).

La protéine responsable de ce transport, appelée tRip, a été identifiée par homologie de séquence avec la protéine Trbp111 (Morales, *et al*., 1999), considérée comme le module de liaison aux ARNt universel. Ce domaine de liaison à l'ARNt est situé dans le domaine C-terminal de tRip, il est exposé à la surface du parasite, et est relié au domaine N-terminal interne par une région transmembranaire (Figure 52E). tRip n'est pas essentielle au développement du parasite, cependant son absence ralentit le développement du stade sanguin (le développement du parasite KO, qui n'exprime pas le tRip, est plus lent dans le sang que celui du parasite sauvage). De plus, l'étude comparative des protéomes des deux parasites (données non publiées) montre qu'en absence de tRip, la synthèse de certaines protéines est affectée. Parmi elles, des protéines impliquées dans l'invasion (MSP1), la prolifération (SEC13) et dans la virulence (SBP1) du parasite, sont sous exprimées dans le parasite KO, ce qui pourrait expliquer l'inhibition de son développement.

**Source et disponibilité des ARNt dans le stade sanguin**

Mon équipe a démontré l'import d'ARNt exogènes au stade infectieux du parasite, le sporozoïte (Bour, *et al*., 2016). Le sporozoïte a été initialement choisi car c'est un stade extracellulaire du parasite, capable d'effectuer la traversée cellulaire. Cette traversée cellulaire est l'un des rares moments du cycle où le parasite est en contact direct avec le cytoplasme de la cellule hôte et donc avec ses ARNt. Cependant, de façon surprenante, le phénotype du parasite KO n'est détectable qu'aux stades sanguins *in vivo* (Bour, *et al*., 2016). Une première hypothèse pour corréler l'import des ARNt de l'hôte avec ce phénotype serait que seul le sporozoïte importe les ARNt, mais que ceux-ci sont utilisés plus tard au cours du stade sanguin. Cette option semble peu probable, car les ARNt importés seraient alors fortement dilués lors de la prolifération du sporozoïte dans les hépatocytes. Une deuxième possibilité serait que les ARNt importés par les sporozoïtes activent immédiatement une voie spécifique, qui n'a pas d'incidence sur le développement

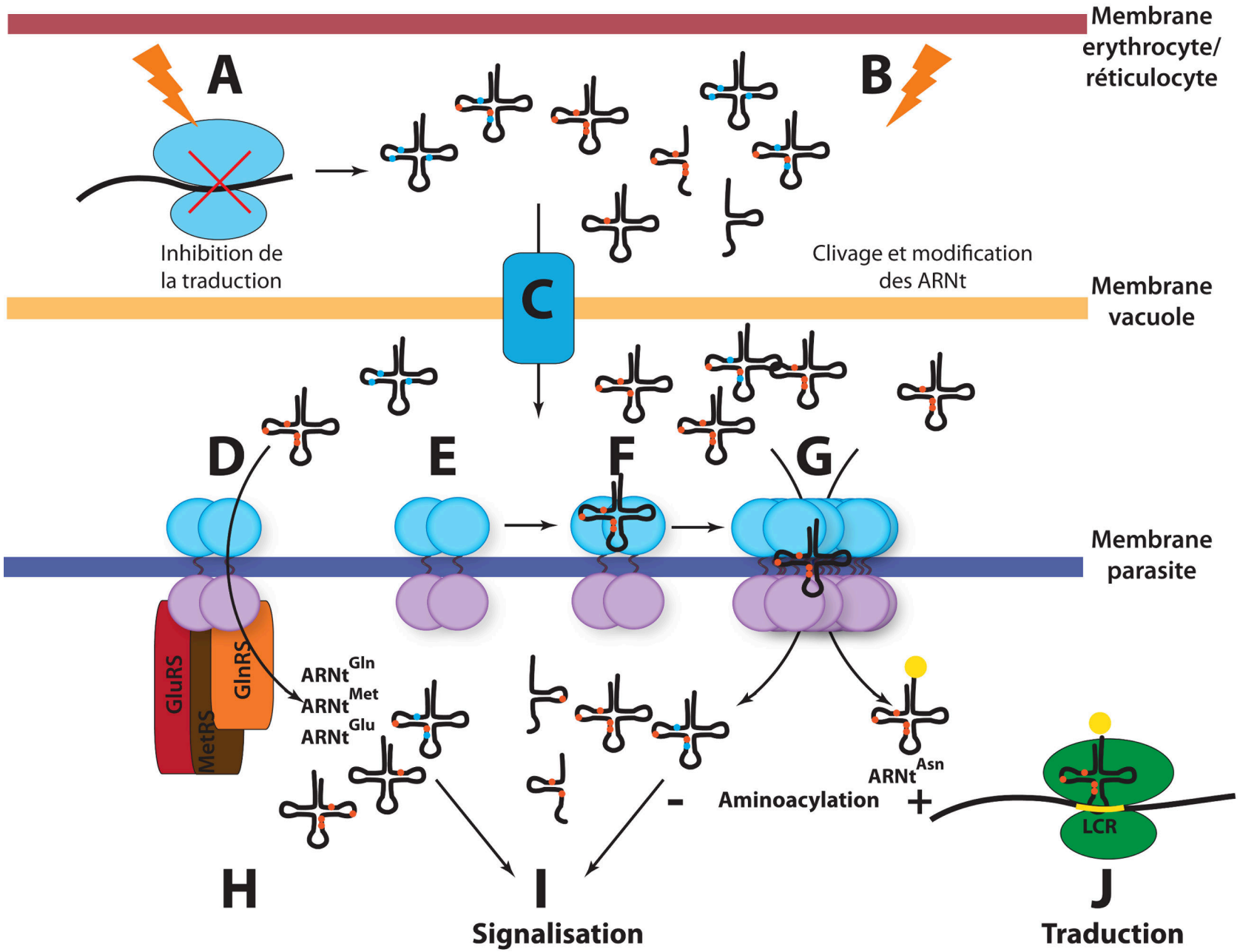

Membrane erythrocyte/ réticulocyte
A
B
Inhibition de la traduction
Clivage et modification des ARNt
C
Membrane vacuole
D
E
F
G
Membrane parasite
GluRS
MetRS
GlnRS
ARNt Gln
ARNt Met
ARNt Glu
ARNt Asn
- Aminoacylation +
LCR
H
I
J
Signalisation
Traduction

**Figure 52 : Récapitulatif de mes recherches et celles de l'équipe sur l'import des ARNt dans la parasite du paludisme.** L'invasion du parasite produit un stress dans la cellule hôte qui peut induire **(A)** l'inhibition de la traduction et/ou **(B)** un changement au niveau des profils de modifications postranscriptionnelles des ARNt. Les modifications favorisant (orange) et défavorisant (bleu) l'interaction des ARNt avec tRip sont représentées. **(C)** Un transporteur localisé dans la membrane de la vacuole parasitophore pourrait permettre le passage des ARNt de l'hôte dans la vacuole parasitophore. **(D)**, **(E)**, **(F)**, **(G)** représentent les différentes formes de tRip identifiées à ce jour (*in vitro*). Le domaine C-terminal extracellulaire, impliqué dans la reconnaissance des ARNt, est représenté en bleu, alors que le domaine N terminal intracellulaire qui interagit avec les GluRS (rouge), GlnRS (orange) et MetRS (marron) est en mauve. Une fois à l'intérieur du parasite, les différents ARNt isodécodeurs Glu, Gln et Met pourraient être immédiatement pris en charge (ou pas) par leurs aaRS homologues **(H)** ; certains ARNt ne seront pas aminoacylés et pourront alors participer à la signalisation dans le parasite **(I)** alors que d'autres seront de bons substrats pour les aaRS du parasite et seront alors intégrés dans la traduction **(J)**, comme il semble que ce soit le cas pour l'ARNt$^{Asn}$.

du sporozoïte, mais qui régule le développement ultérieur des stades sanguins. Enfin, la dernière hypothèse, est celle que j'ai étudiée : les stades sanguins importeraient eux-mêmes des ARNt. Mes résultats préliminaires semblent indiquer que c'est effectivement le cas, puisque je détecte des ARNt de l'hôte dans le parasite sauvage et pas dans le parasite KO (contrôle négatif). Cependant, dans cette hypothèse, la disponibilité des ARNt dans l'érythrocyte et la présence d'une vacuole parasitophore sont des arguments qui vont à l'encontre de cet import.

En effet, aux stades sanguins, les ARNt du globule rouge doivent traverser la membrane de la vacuole parasitophore pour atteindre tRip localisée à la surface du parasite. Comme *Plasmodium* modifie la membrane de la vacuole parasitophore pour l'adapter à ses besoins (Revu dans Spielman et Gilberger, 2009), il est possible que le parasite synthétise une protéine (spécifique ou pas) facilitant le passage des ARNt à travers la membrane de la vacuole (Figure 52C).

De plus, les globules rouges matures contiennent 100 fois moins d'ARN total que les hépatocytes. Ce paradoxe pourrait expliquer la grande préférence que *P. berghei* a pour les réticulocytes (avec un affinité 150 fois supérieure par rapport aux erythocytes matures (Cromer, *et al.*, 2006)). Les réticulocytes proviennent de la moelle osseuse et sont libérées dans le sang, où ils se transforment rapidement en globules rouges. Chez la souris, le nombre de réticulocytes varie entre 1 et 6% des cellules sanguines totales (Bhardwaj et Saxena, 2013). Contrairement aux érythrocytes, les réticulocytes sont des cellules avec une traduction très active, donc riches en ARNt. Cependant, dans une cellule très active en traduction, les ARNt

sont associés soit à leur synthétase soit au facteur d'élongation, soit au ribosome, et ne sont donc pas forcément accessibles pour être importés dans le parasite.

**Le stress cellulaire et ses effets sur la traduction**

Lors de l'infection, le parasite génère une situation de stress intense dans la cellule hôte. Or, des stress, tels que l'exposition à la chaleur, aux métaux lourds, aux agents oxydants, aux concentrations élevées en sels, à l'hypoxie ou au manque de nutriments, provoquent l'inhibition de la traduction. Dans les cellules eucaryotes, le mécanisme le mieux étudié est la phosphorylation du facteur d'élongation eIF2α (Sonenberg et Hinnebusch, 2009) qui bloque la synthèse protéique (Figure 2A). De même, les ARNt sont impliqués dans la réponse au stress (Figure 52B). En effet, suite à un stress, tel que le manque de nutriments, des fragments d'ARNt ont été mis en évidence (Levitz, *et al.*, 1990). Ces fragments ont été identifiés dans tous les organismes étudiés, tels que l'Homme, les plantes et *S. cerevisiae* (Revu dans Thompson et Parker, 2009). Ce clivage se produit préférentiellement dans la boucle anticodon (Haiser, *et al.*, 2008), il a lieu également dans d'autres boucles (Kawaji, *et al.*, 2008) et n'est pas spécifique d'un ARNt donné (Fu, *et al.*, 2009). Par exemple, les fragments 5' provenant des $ARNt^{Ala}$ et $ARNt^{Cys}$ de vertébrés sont responsables de l'inhibition de l'initiation de la traduction en interférant avec l'assemblage du complexe eIF4F (Ivanov, *et al.,* 2011). Il faut noter que les rôles moléculaires précis de la majorité des fragments d'ARNt 5' et 3', que leur localisation, leur abondance relative et leur stabilité restent à déterminer, mais il a été montré que ce processus de clivage n'affecte pas significativement la quantité d'ARNt entiers dans la cellule (Masaki et Ogawa, 2002).

On peut donc raisonnablement imaginer que l'infection des cellules hôte par le parasite provoque une situation de stress, à laquelle la cellule répond en inhibant sa traduction, libérant ainsi les ARNt de leurs partenaires moléculaires et augmentant la concentration d'ARNt libres et accessibles pour être importés dans le parasite (Figure 52A et B).

**Un mécanisme d'import particulièrement efficace et spécifique ?**

Au laboratoire, les travaux du Dr Anne Théobald-Dietrich suggèrent un fonctionnement très particulier de tRip au contact des ARNt. *In vitro*, lorsque tRip est seule, elle forme un homodimère (Figure 52E), cet homodimère se lie dans un premier temps efficacement aux ARNt (Figure 52F), mais rapidement la protéine oligomérise (Figure 52G) et protège complètement l'ARNt des RNAses, suggérant que tRip forme une sorte de tunnel dans lequel l'ARNt est entièrement englobé. Au niveau du parasite, on peut donc imaginer que le pore n'est

pas préformé, mais que les dimères de tRip sont localisés de façon homogène à la surface du parasite, augmentant ainsi la probabilité d'interagir avec les ARNt disponibles, et que l'interaction avec un ARNt va déclencher immédiatement la formation d'un pore dans la membrane qui ne sera alors perméable qu'aux ARNt.

**Reconnaissance et import préférentiel des ARNt**

Le domaine C-terminal de tRip reconnait la structure tertiaire des ARNt (Bour, *et al*., 2016). Cependant un des buts de ma thèse était d'étudier l'affinité de tRip pour les différentes espèces d'ARNt afin de voir si tous sont reconnus et donc potentiellement importés dans le parasite avec la même efficacité. Grâce à la technique du MIST, j'ai montré que l'affinité de tRip varie en fonction de chaque isoacepteur et même de chaque isodécodeur dans certains cas. Cette différence d'affinité est dirigée par les modifications post-transcriptionnelles de chaque ARNt. Selon mes observations, certaines modifications telles qu'une N4-acétylcytidine ($ac^4c$) en position 12, une méthylguanosine (Gm) en position 18 et une pseudouridine (Ψ) aux positions 27 et/ou 28 favoriseraient l'affinité de tRip. Au contraire, une $\Psi_{13}$ et une 5-méthylcytidine ($m^5C$) en position 50 diminueraient cette affinité. Ce « code » de reconnaissance semble particulièrement complexe et est sans précédent puisqu'à notre connaissance il n'existe pas d'autre exemple de protéines capable de discriminer des jeux de bases modifiées. Face à ces difficultés, nous envisageons d'utiliser des ARNt natifs de *Saccharomyces cerevisiae* et/ou d'*Escherichia coli* (déjà partiellement purifiés) pour confirmer mes résultats. Ces expériences supplémentaires permettraient de déterminer si la nature de la modification est importante ou si seule sa position (quelle que soit la modification) constitue un signal pour tRip et de comprendre si les effets de chaque modification se cumulent et ainsi modulent l'affinité de tRip pour les ARNt de l'hôte.

**Le stress cellulaire et ses effets sur les modifications post transcriptionnelles des ARNt**

Des études récentes indiquent que les modifications post-transcriptionnelles des ARNt sont impliquées dans les réponses cellulaires au stress (Revu dans Thompson et Parker, 2009). L'altération des modification post-transcriptionnelles agissent directement sur la stabilité et la fonction des ARNt (Alexandrov, *et al*., 2006, Motorin et Helm, 2010) et en conséquence pourraient avoir un impact direct sur la sélection des codons et le résultat de la traduction (e.g. Begley, *et al*., 2007, Chan, *et al*., 2012, Emilsson, *et al*., 1992). Par exemple, dans les cellules humaines (HeLa), l'enzyme ALKBH1 est surexprimée lorsque la concentration en glucose est faible. Cette enzyme reconnaît spécifiquement le nucléotide modifié $m^1A_{58}$ dans la boucle T des

ARNt et hydrolyse la liaison entre le méthyl et le nucléoside (Liu, *et al.*, 2016). Comme cette modification est essentielle pour la stabilité de l'ARNt$^{Met}$ initiateur chez les levures et chez l'homme (Anderson, *et al.*, 1998, Liu, *et al.*, 2016), ALKBH1 inhibe l'initiation de la traduction. Dans ce contexte, le parasite en augmentant le stress dans la cellule hôte lors de l'invasion, peut agir sur le profil des modifications post-transcriptionnelles des ARNt et donc diriger directement la nature des ARNt qui seront importés (Figure 52B).

**Le rôle des ARNt importés dans le parasite**

La fonction de ces ARNt de l'hôte une fois dans le cytoplasme du parasite est encore inconnue. La caractérisation du parasite KO a montré que la synthèse protéique est réduite dans le stade sanguin (Bour, *et al.,* 2016) et la comparaison des protéomes de parasites KO et sauvage, indique que l'expression d'un nombre limité de protéines est affectée. Ces observations vont dans le sens qu'au moins une partie des ARNt importés sont impliqués dans l'expression des gènes, que ce soit en agissant directement sur la traduction ou que ce soit en agissant sur la transcription ou la stabilité des ARN messagers (Figure 52I et J). La fonction des ARNt importés va dépendre essentiellement de leur capacité à être ou pas aminoacylés par les aminoacyl-ARNt synthétases du parasite. Si l'ARNt de l'hôte est un substrat pour l'aaRS homologue du parasite, celui-ci participera à la synthèse des protéines du parasite (Figure 52 J). En revanche, si l'ARNt de l'hôte n'est pas aminoacylé, il est disponible pour effectuer d'autres fonctions (Figure 52I).

La comparaison des deux protéomes a mis en évidence qu'une majeure partie des protéines sous-exprimées dans le parasite KO sont des protéines de fonction inconnue. Ces protéines ont cependant en commun la présence d'insertions particulièrement riches en asparagine (LCR). Or, la traduction de si longues répétitions d'asn nécessite une grande quantité d'Asn-ARNt$^{Asn}$ dans le parasite (Frugier, *et al*., 2010). Mes résultats indiquent que l'ARNt$^{Asn}{}_{GTT}$ est l'un des 4 ARNt le mieux reconnus par tRip, et serait donc parmi les ARNt préférentiellement importés. En l'absence d'import, la raréfaction d'Asn-ARNt$^{Asn}$ pourrait stopper complètement le décodage du ribosome et inhiber la traduction. Ce processus serait comparable à ce qu'il se passe dans les cellules métastatiques du cancer du sein, où l'augmentation des concentrations en ARNt$^{Glu}{}_{CUU}$ et $^{Arg}{}_{GCC}$ améliore la stabilité et la traduction des ARNm enrichis en codons correspondants et conduit à une augmentation spécifique de la production de protéines métastatiques (Goodarzi, *et al.*, 2016).

Des études en cours dans le laboratoire (Delphine Kapps et José Refugio Jaramillo) indiquent que dans le stade sanguins, le domaine N-terminal de tRip, localisé à l'intérieur du parasite, interagit avec trois partenaires protéiques : la glutamyl-ARNt synthetase (GluRS), la glutaminyl-ARNt synthetase (GlnRS), et la méthionyl-ARNt synthetase (MetRS). D'après mes résultats, les $ARNt^{Met}$ et $ARNt^{Glu}$ sont bien reconnus par tRip (les sondes utilisées n'ont pas permis d'analyser l'affinité de l'$ARNt^{Gln}$). Il est donc possible que certains ARNt isodécodeurs Met, Glu et Gln de l'hôte jouent un rôle particulier dans le développement du parasite (Figure 52H).

Enfin, en dehors de leur rôle canonique dans la traduction, les ARNt sont aussi impliqués dans une grande variété de voies de signalisation. Les revues Kirchner et Ignatova, 2016 et Schimmel, 2017 rendent compte de ces «nouvelles» fonctions, la plupart étant liées à la réponse au stress. Chez *Staphyloccocus aureus*, les protéines Mpf1 et Mpf2 (Multiple peptide resistance), utilisent respectivement des Ala-$ARNt^{Ala}$ et $LysARNt^{Lys}$ pour modifier les lipides de la membrane de la bactérie et lui permettre de résister aux antibiotiques cationiques (Roy et Ibba, 2008). Chez les plantes, les archaebactéries et les bactéries, l'$ARNt^{Glu}$ est impliqué dans la synthèse du précurseur (ALA) de la chlorophylle et de l'hème (Revu dans Jahn, *et al*., 1992). Les ARNt de l'hôte importés dans le parasite, s'ils ne sont pas aminoacylés, ne seront pas pris en charge par la machinerie de traduction du parasite. On peut imaginer que ces ARNt entiers ou clivés puissent alors entrer dans une voie de signalisation (Figure 52I). Dans ce contexte, une de nos hypothèses favorites serait qu'en important les ARNt des différents types cellulaires traversés (chaque type cellulaire exprimant un profil d'ARNt spécifique), le sporozoïte pourrait se « repérer » et reconnaître l'environnement cellulaire optimal pour son développement grâce à la présence d'un ou plusieurs ARNt isodécodeurs spécifiques des hépatocytes. Cette hypothèse est supportée par le fait que le parasite KO traverse plus de cellules, plus longtemps que le parasite sauvage *in vitro*, suggérant qu'un signal lui indiquant de s'installer dans une cellule n'est pas disponible dans le parasite KO.

# MATERIEL ET METHODES

# I. MATERIEL

## 1. Produits chimiques

| | |
|---|---|
| [$^{32}$P]-PP | Perkin Elmer |
| Acétate de magnésium | VWR |
| Acétate de plomb | Suprapur |
| Acétate de potassium | Merck Millipore |
| Acétate de sodium | Carl Roth |
| Acide acétique | Roth |
| Acide chlorydrique | AnalaR NORMAPUR |
| Acide trichloroacétique (TCA) | Carl Roth |
| Acrylamyde/bisacrylamyde (Rotiphorèse Gel 30/40) | Carl Roth |
| Ampicilline | Sigma-Aldrich |
| Arabinose | Sigma-Aldrich |
| ß-Mercaptoéthanol | Carl Roth |
| Bicarbonate de potassium | Sigma-Aldrich |
| Bicarbonate de sodium ($NaHCO_3$) | Sigma-Aldrich |
| Bleu de Coomassie (Instant Blue) | Expedeon |
| Bromure d'éthidium | ICN |
| Chloramphenicol | Amersham Pharmacia |
| Chlorophorme | Carl Roth |
| Chlorure d'ammonium ($NH_4Cl$) | Sigma-Aldrich |
| Chlorure de calcium ($CaCl_2$) | Roth |
| Chlorure de magnesium ($MgC_2$) | VWR |
| Chlorure de potassium (KCl) | Panreac |
| Chlorure de sodium (NaCl) | VWR |
| Chlorure de zinc ($ZnCl_2$) | Sigma-Aldrich |
| Citrate de sodium | Riedel-deHaën |
| DMEM | GIBCO |
| DDM | CliniSciences |
| DTE | Sigma-Aldrich |
| EDTA | Euromedex |
| Éthanol | VWR |
| Glucose | Sigma-Aldrich |
| Glutathion (agarose) | Sigma-Aldrich |
| Glutathion (réduit) | Sigma-Aldrich |
| Glycérol | VWR |
| Glycogène | Qbiogene |
| Héparine | Sigma-Aldrich |
| HEPES | Euromedex |
| Hydroxyde de potassium (KOH) | Sigma-Aldrich |
| Hydroxyde de sodium (NaOH) | Sigma-Aldrich |
| Imidazole | Sigma-Aldrich |
| IPTG | Fisher-Scientific |
| Isofluorane | Piramal Healthcare |

| | |
|---|---|
| Isopropanol | Sigma-Aldrich |
| $K_1$ - Dihydrogénophosphate de potassium ($H_2PO_4^-$, $K^+$) | Merck Millipore |
| $K_2$ - Hydrogénophosphate de potassium ($HPO_4^{2-}$, $2K^+$) | Merck Millipore |
| Kanamicine | Sigma-Aldrich |
| Kétamine | Merial |
| LB | Carl Roth |
| Méthanol | VWR |
| $Na_1$ – Dihydrogénophosphate de sodium ($NaH_2PO_4$) | Merck Millipore |
| $Na_2$ – Hydrogénophosphate de sodium ($Na_2HPO_4$) | Merck Millipore |
| Ocry-gel | TVM |
| Pénicilline-streptomycine (pen-strep) | GIBCO |
| Percoll | GE Healthcare |
| Persulfate d'ammonium (APS) | Carl Roth |
| Phénol | Carl Roth |
| SDS | Sigma-Aldrich |
| Spermidine | Sigma-Aldrich |
| Sulfate d'ammonium | Roth |
| ß-mercaptoethanol | Carl Roth |
| Tampon d'hybridation (UltraSensitive Buffer) | Ambion |
| TBE | Euromedex |
| TEA-Cl | Sigma-Aldrich |
| TEMED | Acros |
| TriReagent | Sigma-Aldrich |
| Tris-base | Sigma-Aldrich |
| Triton | Carl Roth |
| Tween 20 | Sigma-Aldrich |
| Urée | Sigma-Aldrich |
| X-Gal | Euromedex |
| Xylazine | Bayer |

## 2. Nucléotides, oligonucléotides et acides aminés

| | |
|---|---|
| [$\alpha^{32}$P]-ATP | Amersham Pharmacia |
| [$\gamma^{32}$P]-ATP | Amersham Pharmacia |
| ADN de sperme de saumon | Invitrogen |
| ATP, CTP, CMP, UTP, GTP | Sigma-Aldrich |
| L-[$^{14}$C]-tyrosine | Perkin Elmer |
| L-tyrosine | Merck Millipore |
| Oligo dT | Sigma-Aldrich |
| Oligonucléotides | IDT |
| Sondes biotinilées | IDT |

## 3. Acides nucléiques non commerciaux

ARN total de *E. coli :* Produit au laboratoire

$ARNt^{Asp}$ natif de *S. cerevisiae :* Produit au laboratoire selon le protocole développé dans l'article (Dock, 1984).

## 4. Enzymes de restriction et de modification

| | |
|---|---|
| AMV Reverse Transcriptase | Life Sciences |
| BamHI | Thermo Scientific |
| Bstn1 | Thermo Scientific |
| Dynazyne Ext | Thermo Scientific |
| EcoRI | Thermo Scientific |
| FastAP | Thermo Scientific |
| HindIII | Thermo Scientific |
| NsiI | Fermentas |
| RNAse A | Thermo Scientific |
| $S_1$ RNase | Thermo Scientific |
| SphI | Thermo Scientific |
| $T_1$ RNase | Fermentas |
| $T_4$ ADN ligase | Thermo Scientific |
| $T_4$ polynucléotide kinase (PNK) | Thermo Scientific |
| $V_1$ RNase | Ambion |
| Phosphodiestérase de venin de cobra (VDE) | Boehringer Mannheim |

## 5. Enzymes préparées au laboratoire

CCase : nucléotidyl-transférase utilisée pour l'addition d'une séquence CCA en 3' d'un ARNt. Cette enzyme a été produite au laboratoire à partir de la méthode proposée dans l'article Cudny et Deutscher, 1986.

Taq polymérase : enzyme utilisée pour l'amplification d'ADN par PCR, purifiée par Guillaume Bec.

$T_7$ RNA polymérase : enzyme utilisée pour la transcription *in vitro*. Purifiée en suivant le protocole décrit dans l'article Wyatt, *et al.*, 1991.

## 6. Plasmides

pDrive ($Amp^R$, $Kan^R$) : vecteur contenant une origine de transcription pour l'enzyme $T_7$ RNA polymérase avec le promoteur correspondant.

pGEX-2T ($Amp^R$): plasmide qui permet la fusion de la protéine glutathioine-S-transférase (GST) à l'extrémité C-Terminale de la protéine à produire.

pGro7 ($Cam^R$) : ce plasmide contient les gènes codant pour deux chaperonnes (groES et groEL). Il est utilisé pour assurer le repliement correct de la protéine d'intérêt.

pQE30/pQE70 ($Amp^R$) : ces deux plasmides permettent la fusion d'une séquence de six histidines en N-Terminal (pQE30) ou en C-Terminal (pQE70) de la protéine d'intérêt. Ils possèdent aussi l'opérateur lac qui permet de contrôler l'induction de la production de la protéine.

pUC119 ($Amp^R$) : la séquence de ce plasmide est dérivée du vecteur pUC18 (Yanisch-Perron, *et al.*, 1985), et comprend une origine de réplication pBR322 et une origine de réplication du phage F1.

## 7. Cellules pour clonage et production d'ARNt

ER2566 : Souche bactérienne utilisée pour l'expression de *Pf*tRip$_{214-402}$ et *Pf*tRip$_{214-402}$ Mut. Génotype : F- λ- fhuA2 [lon] ompT lacZ::T7 gene 1 gal sulA11 Δ(mcrC-mrr)114::IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS) endA1 [dcm].

TG1 : souche bactérienne utilisée pour la production d'ADN plasmidique codant pour les ARNt. Génotype : F'[traD36 pro $AB^+$ $lacI^q$ lacZΔM15] supE Δ(hsdM-mcrB)5 thi Δ(lac-proAB).

Top10 : Souche bactérienne utilisée pour l'expression des protéines et le clonage des aptamères. Génotype : F mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZ ΔM15 ΔlacX74 deoR recA1 araD139, Δ(ara-leu)7697 galU galKrpsL endA1 nupG.

HeLa : Lignée cellulaire humaine utilisée pour la purification d'ARN de transfert produits *in vivo*.

## II. METHODES

## 1. Préparation des parasites

Le parasite *Plasmodium berghei* a été utilisé tout au long de cette thèse. J'ai utilisé une souche sauvage ainsi qu'une souche mutante n'exprimant pas la protéine tRip (*Pb*tRip-KO) pour les études *in vivo*. Cependant, la culture *in vitro* de cette espèce de *Plasmodium* n'est pas réalisable. Il faut donc reconstituer son cycle naturel *en laboratoire*. Dans cette partie, je répertorie les différentes souches de *P. berghei* utilisées, et son mode de production.

### *1.1. Souches de Plasmodium utilisées dans cette étude*

*Pb* GOMO14 : souche de *P. berghei* exprimant constitutivement la protéine GFP sous le contrôle du promoteur de la protéine hsp70 (Manzoni, *et al.*, 2014). L'intensité de fluorescence de ce parasite est forte.

*Pb*tRip-KO : souche de *P. berghei* dans laquelle le gène codant pour tRip a été remplacé par le gène codant pour la protéine m-Cherry (Bour, *et al.*, 2016). Plusieurs clones ont étés produits (V3, V4, V5 et R7), mais les clones utilisés au cours de ma thèse sont essentiellement les clones V4 et V3.

Ces parasites ont été produits comme indiqué dans la Figure 53 et les paragraphes *1. 2.* et *1. 3.*. Les parasites produits ont été utilisés soit pour purifier des ARN totaux de *Plasmodium*, soit pour faire des stocks de parasites.

### *1.2. Infection des souris*

Les souris utilisées sont des souris Swiss, achetées à l'âge de trois semaines chez Envigo. Elles sont mises en cage le jour de leur arrivée à l'animalerie, avec un apport d'eau et de nourriture continu, avec un changement de litière une fois par semaine

Contrairement au parasite *Pb* GOMO14, le parasite *Pb*tRip-KO a une infectivité réduite (Bour, *et al.*, 2016), et se développe de façon aléatoire dans les souris adultes. C'est pour cette raison que nous utilisons de jeunes souris dans nos expériences. L'âge minimal des souris est fixé à quatre semaines, pour respecter la semaine de quarantaine après arrivage à l'animalerie conseillée par l'éthique et l'âge maximal est fixé à six semaines, âge à partir duquel les souris ont un système immunitaire mature et sont donc considérées comme adultes.

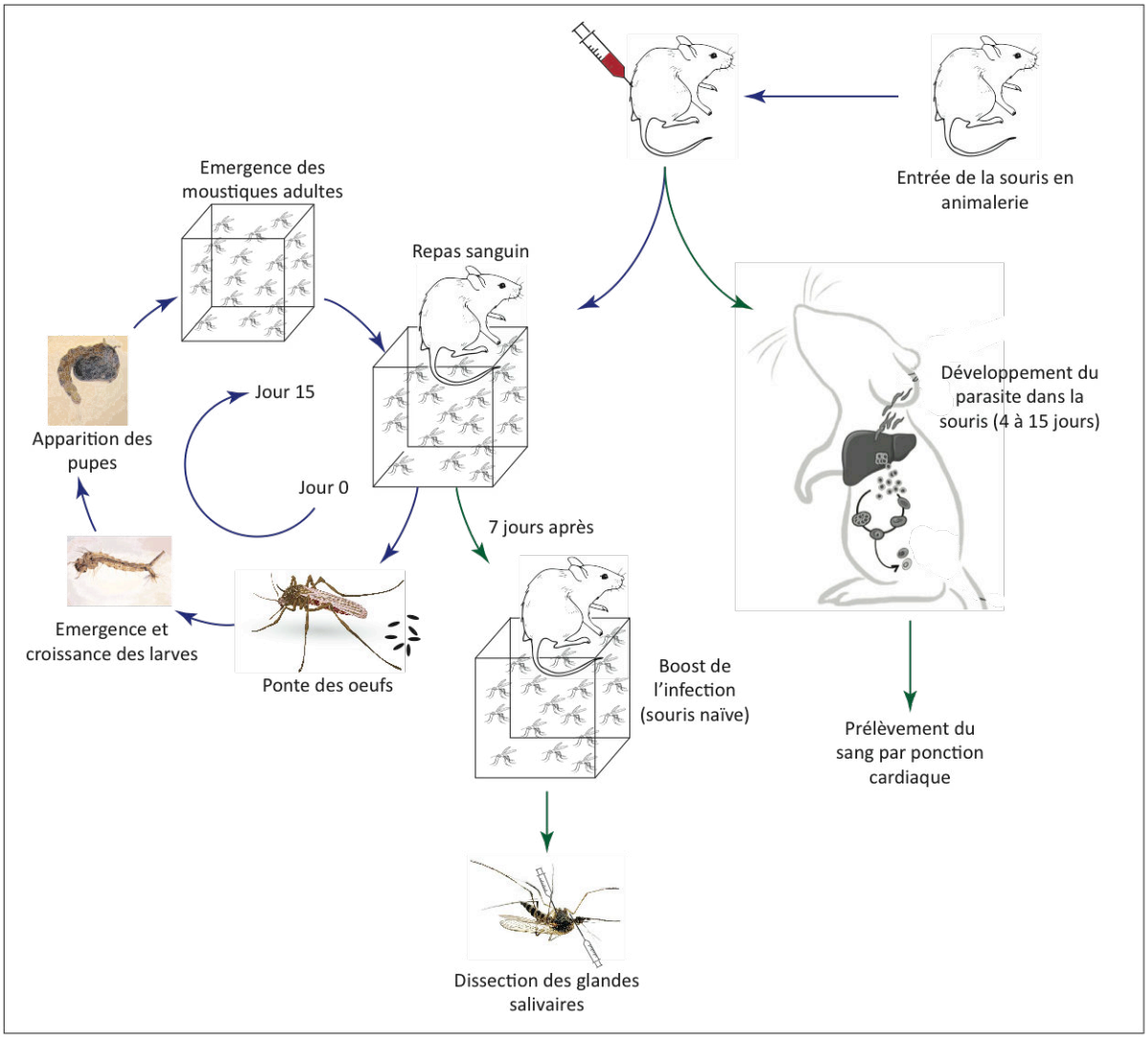



**Figure 53 : Maintient de la lignée des moustiques (flèches bleues) et production du parasite *Plasmodium berghei* (flèches vertes).**

*a. Inoculation des parasites*

L'infection des souris par *P. berghei* peut se faire de deux façons : avec des sporozoïtes ou avec des stocks de sang de souris parasitées.

Les sporozoïtes sont récupérés après dissection des glandes salivaires de moustiques infectés, et environ 10000 sporozoïtes dilués dans un volume maximal de 150 µL de milieu DMEM sont injectés dans une souris par injection intra-veineuse (i.v.). Avec cette méthode, les parasites sont détectés dans le sang entre 7 et 10 jours après l'injection.

Alternativement, 100 à 200 µL de sang parasité par *Pb* GOMO14, et 250 à 400 µL de sang parasité par *Pb*tRip-KO, sont injectés dans la souris en intra-péritonéale (i.p.). Dans ce cas, la détection des parasites dans le sang est possible entre 4 et 6 jours après l'inoculation.

*b. Suivi de l'infection*

Dans les deux cas, le développement du parasite dans le sang des souris (parasitémie) est suivi quotidiennement. Pour cela, une goutte de sang est prélevée au niveau de la queue de la souris, et est étalée sur une lame de microscopie. Le frottis est séché à l'air, fixé au méthanol 100 %, coloré (Diff-Quick, Sigma-Aldrich) et observé au microscope (objectif X100). Les érythrocytes totaux et les érythrocytes parasités sont comptés dans trois à cinq champs par frottis, pour calculer le pourcentage de globules rouges infectés par rapport au nombre total de globules rouges. Cette parasitémie doit atteindre des niveaux différents selon l'utilisation ultérieure de la souris infectée :

- 2 à 5 % pour infecter des moustiques,
- environ 10 % pour préparer des stocks de sang parasité,
- au minimum 10 % en évitant cependant de faire souffrir les souris pour la purification des ARN totaux de *Plasmodium* à partir des stades sanguins.

Une fois la parasitémie désirée atteinte, la souris est anesthésiée soit par injection de 100 µL d'anesthésique liquide (PBS 1 x, 10 mg/mL xylazine, 100 mg/mL kétamine) par gramme de souris, soit par anesthésique gazeux (isofluorane). L'efficacité de l'anesthésie est vérifiée par l'absence de reflexes.

*c. Prélèvement du sang parasité par ponction cardiaque*

La ponction cardiaque est la technique qui permet le prélèvement d'un volume maximal de sang, soit environ 1 mL de sang pour une souris de 25 à 30 g. Une fois la souris endormie, elle est mise sur le dos, et les pattes avant sont bien écartées. A l'aide d'une aiguille (0,45 x 13 mm) et d'une seringue de 1 mL héparinisée, le sang est aspiré doucement directement à partir du coeur. La souris doit rester vivante puisque ce sont les battements de cœur qui propulsent le sang à l'intérieur de la seringue.

*d. Stocks de parasite au stade sanguin*

Plusieurs types de stocks sont disponibles au laboratoire. Les parasites peuvent être amplifiés par passages successifs d'une souris à une autre sans passer par les stades de développement dans le moustique. Pour ce projet, nous limitons le nombre de passages à 3,

pour éviter une dérive génétique du parasite qui aurait pour conséquence la délétion de gènes utilisés au cours de son développement dans l'hôte invertébré.

P0 : Sang de souris infectées directement par des sporozoïtes extraits des glandes salivaires de moustiques infectés.

P1 : Sang de souris infectées par un stock P0 (passage 1).

P2 : Sang de souris infectées par un stock P1 (passage 2).

Le sang de souris infecté est prélevé par ponction cardiaque et conservé à -80°C dans 2 volumes d'Alsever's solution (Sigma-Aldrich) complémenté avec 11% glycérol.

### *1.3. Les moustiques*

Des moustiques *Anopheles stephensi* ont étés utilisés parce que cette espèce permet d'atteindre des charges parasitaires très élevées (jusqu'à 100000 sporozoïtes/moustique), ce qui réduit le nombre de moustiques à disséquer.

#### *a. Maintient de la lignée*

Les moustiques ont étés élevés à l'insectarium, à une température de 28ºC, une humidité de 75-80% et un cycle jour/nuit de 12 h/12 h. Un repas sanguin est nécessaire pour permettre aux femelles de pondre. Une souris naïve est anesthésiée, et posée sur la cage de moustiques, pendant environ 20 min, pour permettre aux moustiques femelles de se nourrir de son sang. Les yeux de la souris sont protégés par dépôt de gel oculaire (Ocry-gel). Un petit coussin chauffé est posé sur la souris pour la protéger de l'hypothermie, et pour accélérer le repas sanguin en attirant les moustiques vers la source de chaleur. De plus, la cage est protégée de la lumière par une toile opaque. Ceci protège les yeux de la souris et favorise le repas, puisque les moustiques piquent normalement à la tombée de la nuit. Après le repas sanguin, la cage est remise à 28ºC avec une humidité de 75-80%.

Trois jours plus tard, un pondoir est mis à disposition dans la cage des moustiques. Le lendemain, les œufs récoltés sont mis à éclore dans un bac contenant de l'eau déionisée. Après éclosion, les larves sont nourries matin et soir avec de la nourriture pour poisson finement broyée (TetraMin). Tout le long de leur développement, les larves sont diluées pour optimiser leur croissance, et la quantité de nourriture est adaptée à leur densité et à leur stade de développement.

Neuf à dix jours après la ponte, les premières pupes apparaissent. Elles sont récoltées et placées dans un récipient d'eau propre à l'intérieur d'une cage, pour que les adultes émergent environ 24 h après. Un coton imbibé d'une solution de saccharose 10% (diluée dans de l'eau déionisée) est mis sur la cage (et réhumidifié régulièrement) pour nourrir les moustiques.

*b. Infection des moustiques*

Pour la production de sporozoïtes, une souris impaludée (parasitémie de 2 à 5%) est utilisée pour le repas sanguin. La cage de moustiques est ensuite gardée à 21ºC avec une humidité entre 75 et 80%, et avec un cycle jour/nuit de 12 h/12 h. Un deuxième repas sanguin a lieu une semaine plus tard sur une souris naïve afin de stimuler le développement du parasite. Entre 18 et 21 jours après le repas infectieux, les moustiques sont disséqués.

*c. Dissection des glandes salivaires*

Les moustiques sont aspirés, endormis avec du $CO_2$ puis transférés dans un Falcon de 50 mL. Le tube (bien fermé) est secoué pour assommer les moustiques et permettre l'ajout d'éthanol 70%. Les moustiques sont imbibés pendant 10 à 20 secondes, ce qui les endort et permet de perméabiliser leur cuticule. L'éthanol est lavé avec quatre bains de milieu DMEM supplémenté de 2% d'antibiotiques pénicilline-streptomycine (pen-strep). Les moustiques sont disposés sur des lames de microscopie (en deux lignes de 15 moustiques par lame) et conservés sur la glace jusqu'à leur dissection. Du milieu DMEM 2% pen-strep est régulièrement rajouté sur les moustiques pour éviter leur déshydratation.

La dissection des glandes salivaires est faite sous loupe binoculaire (Zeiss Stemi™ 2000/C), à l'aide de deux aiguilles : l'une maintient le moustique au niveau du thorax sur la lame, alors que l'autre est utilisée pour tirer sur la tête du moustique (Figure ). Ce geste, renforcé par une légère pression sur le thorax du moustique permet de séparer la tête du thorax, et de faire sortir les glandes salivaires qui sont accrochées à la tête du moustique. Les glandes salivaires sont alors séparées de la tête, récoltées avec une pince et placées dans du DMEM 2% pen-strep dans la glace. Elles sont transférées dans un potter à piston (verre borosilicaté de 5 $cm^3$) et broyées. Afin d'éliminer les plus gros fragments provenant du moustiques, le broyat est filtré sur un tamis de diamètre 45 µm et les sporozoïtes sont quantifiés sous microscope (Nikon Diaphot 300) par comptage sur une lame de Kova .

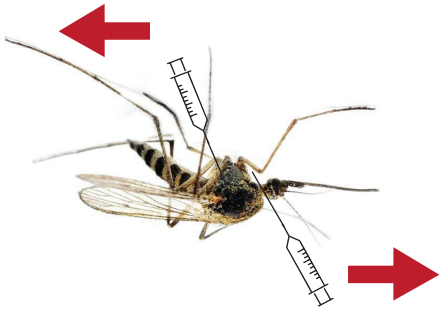


**Figure 54 : Comment disséquer les glandes salivaires du moustique.**

## 2. Préparation des protéines recombinantes

Deux protéines recombinantes de *Plasmodium falciparum* ont été étudiées : *Pf*tRip et *Pfapi*TyrRS. Dans cette partie, la production et purification de ces deux protéines sont détaillées.

### *2.1. PftRip (tRNA import protein de Plasmodium falciparum)*

La protéine *Pf*tRip contient 402 acides aminés. C'est une protéine membranaire, présente uniquement chez les Apicomplexes et qui permet l'entrée des ARNt exogène dans le parasite au stade sporozoïte (Bour, *et al*., 2016). Au cours de cette thèse, trois formes de *Pf*tRip ont été utilisées (Figure 55) : la protéine entière (*Pf*tRip), le domaine C-Terminal (*Pf*tRip$_{214\text{-}402}$) sauvage et le domaine C-terminal muté (*Pf*tRip$_{214\text{-}402}$Mut) fusionnés à la GST. Dans *Pf*tRip$_{214\text{-}402}$Mut, les résidus sérine 312 et méthionine 315 localisés dans le site de liaison aux ARNt ont été substitués par des résidus alanine. Ce mutant ne lie plus les ARNt *in vitro*.

La séquence du gène codant pour *Pf*tRip a été « harmonisée », c'est à dire que les codons rares dans *P. falciparum* ont été remplacés par des codons rares dans *E. coli*. Cette harmonisation des codons permet en théorie de conserver la vitesse de traduction de l'organisme d'origine (*Plasmodium*) pour la surexpression de la protéine recombinante dans un organisme hétérologue (bactérie). Dans le cas de *Pf*tRip, cette étape est essentielle car elle a permis d'obtenir un repliement optimal de la protéine dans *E. coli*. Le gène « harmonisé » de *Pf*tRip a été cloné dans le plasmide pQE30 (Qiagen) et transformé dans des cellules Top10 afin de produire une protéine recombinante fusionnée à un tag 6-histidine (6-His) à son extrémité N-terminale.

```
     10         20         30         40         50         60
MCVLTLVKDD IKSDILKLVL DYIKVTVVQD NENVKLPEIC YDKVKIKKKK IYIYILYMFF

        70         80         90        100        110        120
FQHLRTVFKK LIFFFFFFFH TLYIIMMKYK DAHIYLSIYS YFFSLILIYF FFFSSFKITL

       130        140        150        160        170        180
QYKNKTYKDL FCTLYALIDI YDCYSELFNE DEGKVSENEE FIFHLASDKY ILKQSDMKHL

       190        200        210        220        230        240
NDLLCEKSYI ISNKHASIVD IFYFCAIHKL LDEMAVKERI EFSYIYRWYL HIQETLLANF

       250        260        270        280        290        300
STLKKLIVKD SLENLLNNKT TNNAPEHKNN FVSKESKENK SQNNESPKNK KKDVQNKNNA

       310        320        330        340        350        360
PNKKVEETKK LDDISRLNVL VGYVEQVEIH PDADTLYCLK INLGEDKPRD ICSGLRNKKN

       370        380        390        400        410        420
AEDLLNKYVL VLANLKEKSL RGKKSHGMVL CGSFDEKVEL LVPPNGVKIG ERILFHNMDP

       430        440        450        460        470
NVIPDKNLSS DKEKNPFFHI QPHLILKDGV AHYKDTKWIS SQGDITCVLN QGTIS
```

**Figure 55 : Séquence de la protéine *Pf*tRip.** En bleu, la séquence du domaine C-terminal et en rouge, les deux acides aminés mutés dans *Pf*tRip$_{214\text{-}402}$Mut.

Les gènes codant pour *Pf*tRip$_{214\text{-}402}$ et *Pf*tRip$_{214\text{-}402}$ Mut ont été insérés dans le plasmide pGEX-2T entre les sites *Bam*HI et *Eco*RI, pour produire une protéine de fusion avec la GST en N-Terminal. Ces protéines ont été exprimées dans des cellules ER2566, en présence du plasmide pGro7 qui code pour deux chaperonnes GroEL et GroES, qui aident au repliement correct des protéines surproduites.

*a. Protocole de purification de PftRip*

Les bactéries transformées sont mises en préculture sur la nuit à 37ºC dans du milieu LB contenant de l'ampicilline (0,1 mg/mL) et du glucose (2%, pour réprimer efficacement le promoteur de transcription). Le lendemain, 5 ml de préculture sont utilisés pour ensemencer 250 ml de culture qui est incubée à 30ºC jusqu'à atteindre une densité optique ($DO_{600}$) de 0,6. Les cellules sont alors centrifugées, reprises dans le milieu d'induction (LB, 0,25 mM IPTG, ampicilline) et incubées sous agitation vigoureuse toute la nuit à 18ºC. Les cellules sont centrifugées, lavées avec du PBS et sont soit stockées à -20°C ou utilisées pour purifier la protéine recombinante.

La protéine *Pf*tRip (contenant un 6-His tag en N-terminal) est purifiée selon le protocole établi dans le laboratoire par Anne Théobald-Dietrich. Les cellules sont resuspendues dans

25 mL de tampon $A_2A_3$ (Tableau 10), soniquées (7 min à 140 v dans la glace) et ultracentrifugées pendant 45 min à 45000 rpm.

La première étape consiste à faire une chromatographie d'affinité sur une colonne Ni-NTA (2 mL, BioRad). La colonne est lavée avec 5 mL de tampon $A_2A_3$ (Tableau 10). Un premier gradient de NaCl (100 % $A_1A_2$ à 100 % $B_1$), suivit d'un gradient inverse (100 % $B_1$ à 100 % $A_1$) (Tableau 10) permet d'éliminer les acides nucléiques fixés sur tRip. L'étape de lavage est poursuivie dans la tampon $A_1$ contenant 10 mM d'imidazol afin d'éliminer les protéines fixées de façon aspécifique à la résine. Un gradient d'imidazole (100 % $A_1$ à 100 % $B_2$) permet d'éluer la protéine recombinante. Les fractions correspondant à l'étape d'élution sont analysées sur gel.

Les fractions contenant la protéine sont rassemblées et diluées cinq fois dans du tampon A-UnoQ (Tableau 10). La colonne UnoQ (1 mL, BioRad) est une colonne échangeuse d'anions qui permet de purifier la forme active de la protéine, qui ne reste pas accrochée sur la résine. *Pf*tRip est alors concentrée sur Amicon-Ultra 30K (Merck-Millipore).

La troisième étape consiste à utiliser une colonne Superdex 200 (24 mL, Sigma-Aldrich) dans le but d'éliminer les derniers contaminants et les agrégats de *Pf*tRip par exclusion de taille. *Pf*tRip est éluée en un pic unique après le passage de 15 mL de tampon superdex (Tableau 10), elle est dialysée pendant la nuit à 14ºC dans 250 mL de tampon de dialyse (Tableau 10). Dans ce tampon, *Pf*tRip peut être conservée à -80ºC. Son activité de liaison aux ARNt a été testée avant et après congélation par une expérience de retard sur gel comme indiqué dans les résultats, et ce pour chaque purification.

*b. Protocole de purification des deux formes du domaine C-Terminal de PftRip*

Les bactéries ER2566, ont été transformées avec le plasmide pGro7 (résistance au chloramphenicol) et les plasmide pGEX-2T contenant le gène GFP-*Pf*tRip$_{214-402}$ ou GFP-*Pf*tRip$_{214-402}$ Mut. Des précultures sont incubées à 37ºC pendant la nuit dans 50 mL de milieu LB, glucose 2%, 0,1 mg/mL d'ampicilline et 0,03 mg/mL de chloramphénicol. Le lendemain 5 ml de préculture sont utilisés pour inoculer 250 mL du même milieu de culture
en présence de 0,25 mg/ml d'arabinose. L'arabinose permet d'induire la production des protéines chaperonnes à partir du plasmide pGro7. La culture est incubée à 37ºC sous agitation jusqu'à atteindre une $DO_{600}$ de 2. Les cellules sont centrifugées, resuspendues dans

250 mL de milieu d'induction, contenant du LB, les deux antibiotiques et 0,25 mM d'IPTG, et incubées la nuit à 18ºC sous agitation.

Après centrifugation à 4000 rpm, le culot de cellules est repris dans 25 mL de tampon de charge (phosphate de sodium 50 mM pH 8 et de chlorure de sodium 300 mM). Les cellules sont soniquées (7 min à 120 v) et ultracentrifugées 45 min à 35000 rpm à 4ºC pour enlever les débris cellulaires.

Les protéines ont été purifiées en batch par affinité sur une résine agarose glutathion : 3 mL de résine sont incubés sous agitation pendant 30 min à température ambiante avec l'extrait cellulaire afin de favoriser l'interaction de la GST avec la résine. Le mélange est transféré dans une colonne, et lavé avec 40 mL de tampon de charge. La protéine est alors éluée en incubant la résine 5 min à température ambiante avec 2 mL de tampon d'élution (50 mM phosphate de sodium pH 8, 300 mM chlorure de sodium, 10 mM glutathion réduit). La pureté de la protéine a été vérifié sur gel SDS-PAGE et a finalement été dialysée à 4ºC pendant la nuit contre le tampon de conservation (phosphate de potassium 50 mM pH 7,5, chlorure de potassium 150 mM, et glycérol 50%) et conservée à -20ºC.

### *2.2. Tyrosyl-ARNt synthétase apicoplastique de Plasmodium falciparum : PfapiTyrRS*

Plusieurs formes de la protéine ont été utilisées pour cette étude (Figure 56), en particulier des formes mutantes de la dans lesquelles les 24 acides aminés à l'extrémité N-terminal correspondant au signal de localisation apicoplastique (*Pfapi*TyrRS) ou le domaine C-terminal $S_4$-like (*Pfapi*YRSΔ$S_4$) ont été délétés.

Trois formes du gène codant pour *Pf*apiTyrRS sauvage ont été clonés dans le plasmide pQE30, afin d'obtenir une protéine recombinante avec un Tag 6-His en N-terminal. La séquence de l'ADN était soit la séquence endogène de *Plasmodium falciparum* (Endo) soit une séquence optimisée pour la traduction dans *E. coli* (Opt), soit une séquence harmonisée (Harm) à l'aide du programme ANACONDA (Pinheiro, *et al.*, 2006 ; Moura, *et al.*, 2015) pour récapituler dans *E. coli* les variations de vitesses de traduction rythmées par la présence de codons rares. La conception de ces gènes est expliquée dans l'article Cela, 2018 présenté dans la partie « Résultats » de cette thèse. Les mutants tronqués à l'extrémité N-terminale ont été également clonés dans pQE30, alors que le gène codant pour la protéine *Pfapi*YRSΔ$S_4$ a été clonée dans pQE70, avec une séquence 6-His en C-Terminal.

**Tableau 10 : Composition des tampons utilisés pour la purification de *Pf*tRip**

| Tampons | Cº finale | Composé |
|---|---|---|
| Tampon $A_2A_3$NiNTA | 50 mM<br>300 mM<br>0,1%<br>5 mM<br>10% | HEPES-Na pH8<br>NaCl<br>DDM<br>ßME<br>Glycérol<br>$H_2O$ qsp |
| Tampon $B_1$NiNTA | 50 mM<br>2 M<br>0,005%<br>5 mM<br>10% | HEPES-Na pH8<br>NaCl<br>DDM<br>ßME<br>Glycérol<br>$H_2O$ qsp |
| Tampon $A_1$NiNTA | 50 mM<br>10 mM<br>0,005%<br>5 mM<br>10% | HEPES-Na pH8<br>Imidazole<br>DDM<br>ßME<br>Glycérol<br>$H_2O$ qsp |
| Tampon $B_2$NiNTA | 50 mM<br>300 mM<br>0,005%<br>5 mM<br>10% | HEPES-Na pH8<br>Imidazole<br>DDM<br>ßME<br>Glycérol<br>$H_2O$ qsp |
| Tampon A UnoQ | 50 mM<br>0,005%<br>5 mM<br>10% | HEPES-Na pH8<br>DDM<br>ßME<br>Glycérol<br>$H_2O$ qsp |
| Tampon B UnoQ | 50 mM<br>500 mM<br>10% | HEPES-Na pH8<br>NaCl<br>Glycérol<br>$H_2O$ qsp |
| Tampon Superdex | 25 mM<br>75 mM<br>0,5 mM<br>5 mM<br>5%<br>0,005% | HEPES-KOH pH7<br>KCl<br>$MgCl_2$<br>ßME<br>Glycérol<br>DDM<br>$H_2O$ qsp |
| Tampon Dialyse | 25 mM<br>75 mM<br>0,5mM<br>5 mM<br>5%<br>0,005% | HEPES-KOH pH7<br>KCl<br>$MgCl_2$<br>DTE<br>Glycérol<br>DDM<br>$H_2O$ qsp |

```
        10         20         30         40         50         60
MAPHRLNNII HENNKNNIGT YDIKSKALKK LYERKLIHYV SDIRNIDKIL YHNENEKEKK

        70         80         90        100        110        120
NRKSVYAGID LTCKYLHLGN LVPLITLDIL RNHNTDVIIL LGNSTTQIGD PSFQKVERQK

       130        140        150        160        170        180
TLEKDILENE ENIRRTIIEL FLQREICEED MNELIKKSNI ERDKEFIYES DNKGSLIILK

       190        200        210        220        230        240
NSLWYDKMNI IDFLKYGEYF SINKLLRKEC FLNKIKKNLT LKDLNYITLQ SFDFLHLFNK

       250        260        270        280        290        300
FKTCIQIGGS DQWGNIQSGI ELAQYISNTQ LYGLTTNLLV YKNNIKYSKS QFNENKRLPI

       310        320        330        340        350        360
WIDKNYNSPY LFWNFLRNVE DQKVQSYIDM LTNLKININQ EIETVYNPMD TNLNEILDET

       370        380        390        400        410        420
LDDSNKKKKN NDNNNNNNNV KDINITTEHN NYSNDISLEK SYEEKINQAK KQLSDSVTSY

       430        440        450        460        470        480
IFGEHTVKKI HKMKDVLKNNEFHKINNIDD IKVFPYVEIT MEHINKKQIN ISDLLKKFDI

       490        500        510        520        530
ASTNKEAKEK ISQNCIYLNE LLINDSKYSL NINNFIKLHN NYYAILRLGK RTSYSIIIK
```

**Figure 56 : Séquence de la protéine *Pfapi*TyrRS.** En bleu, la séquence délétée dans le mutant *Pfapi*TyrRS $\Delta S_4$

Les plasmides ont été transformés dans les cellules Top10. Les bactéries transformées sont mises en préculture sur la nuit à 37ºC dans du milieu LB contenant 0,1 mg/mL d'ampicilline et 2% de glucose. Le lendemain, 5 ml de préculture sont utilisés pour ensemencer 250 ml de culture, incubée à 37ºC jusqu'à atteindre une $DO_{600}$ de 0,8. Les cellules sont alors centrifugées, resuspendues dans le milieu d'induction (LB, 1mM IPTG, ampicilline) et incubées sous agitation vigoureuse toute la nuit à 18ºC. Les cellules sont centrifugées, lavées et sont reprises dans 25 mL de tampon (phosphate de sodium 50 mM pH 8 et de chlorure de sodium 300 mM), soniquées (7 min à 140 v dans la glace), ultracentrifugées pendant 45 min à 45000 rpm et le surnageant a été chargé sur une colonne Ni-NTA (2 mL) comme indiqué par le fabricant (Qiagen). La protéine recombinante purifiée est dialysée sur la nuit à 4ºC dans un tampon de conservation (25 mM HEPES-KOH pH 7,5, 25 mM KCl, 50% glycérol) et stockée à -20ºC.

## 3. Préparation des acides nucléiques

Deux types d'ARN de transfert ont été utilisés lors de cette thèse : de l'ARNt natif extrait directement des cellules, et des séquences d'ARNt transcrits *in vitro*. Dans cette partie, la production et la purification des deux types d'ARNt sont détaillées.

### *3.1. Purification d'ARNt natifs à partir de cellules HeLa*

*a. Préparation de l'ARNt total*

Pour chaque purification, un culot contenant entre 80 et 100 millions de cellules HeLa a été utilisé. Les cellules sont resuspendues dans 20 mL de TriReagent. Après une incubation de 5 minutes à température ambiante, 4 mL de chloroforme sont rajoutés, le milieu est vigoureusement vortexé et incubé 15 min à température ambiante. Après centrifugation de 15 min à 11000 g, la phase aqueuse, contenant l'ARN total, est précipitée avec 20 ml d'isopropanol (incubation de 15 min à température ambiante, puis centrifugation de 30 min à 11000 g à 4ºC). Le culot d'ARN est alors repris dans 2 mL d'eau stérile contenant 0,5 mM d'EDTA et filtré sur une colonne Micro Bio-Spin (BioRad) afin d'éliminer le matériel non dissout.

La deuxième étape consiste à purifier les ARNt des autres ARN cellulaires sur une colonne d'exclusion de taille (Superdex 200, 24 mL, Sigma-Aldrich) dans le tampon contenant 50 mM HEPES-KOH pH 7,5 et 150 mM KCl. Les 2 mL ont été injectées en quatre fois sur la colonne, et lavés avec 25 mL de tampon. Seules les fractions contenant les ARNt ont étés précipitées à l'éthanol, puis resuspendues dans de l'eau et quantifiés.

*b. Préparation de la colonne d'affinité*

Nous disposons de quatre sondes biotinilées en 3' (Tableau 11) conçues selon l'article de (Yokogawa, *et al.*, 2010). Chaque sonde s'hybride spécifiquement à la boucle D, le bras D et le bras anticodon des ARNt natifs à purifier. Une dizaine de culots de cellules HeLa correspondant à autant d'extractions d'ARN total ont été utilisés pour obtenir la quantité d'ARNt total nécessaire indiquée dans l'article Yokogawa, *et al.,* (2010) (100 µl à 3,75 µg/µL) pour la purification d'un seul isoaccepteur.

**Tableau 11 : Séquence des sondes d'ADN biotinilées**

| Sondes | Séquences |
|---|---|
| $ARNt^{Ala}$ | 5' CATGCTAAGCACGCGCTCTACCACTGAG-Biotine 3' |
| $ARNt^{Arg}$ | 5' ATCCGTAGTCAGACGCGTTATCCATTGC–Biotine 3' |
| $ARNt^{Asn}$ | 5' TTAACAGCCAAACGCGCTAGCCGATTGC –Biotine 3' |
| $ARNt^{Asp}$ | 5'-CGTGACAGGCGGGGATACTCACCACTATAC –Biotine 3' |

Six nmol de sonde sont fixées sur 200 µL de résine de stréptavidine 100% (Sigma-Aldrich), équilibrée préalablement dans du tampon Tris-HCL 10 mM, pH 7,6 (incubation de 30 min à température ambiante).

*c. Hybridation*

100 µL d'ARNt total à une concentration de 3,75 µg/µL sont ajoutés et mélangés à la sonde fixée sur la résine. 100 µl de tampon d'hybridation, contenant 20 mM Tris-HCl pH 7,6, 1,8 M chlorure de tetraethylammonium (TEA-Cl) et 0,2 mM EDTA, sont ajoutés et mélangés. Le tout est incubé 10 min à 37ºC, puis immédiatement centrifugé 1 min à 750 g afin d'éliminer ce tampon d'hybridation. La résine est lavée quatre fois avec 800 µL de Tris-HCl 10 mM pH 7,6 à température ambiante.

*d. Elution*

400 µL de Tris-HCl 10 mM pH 7,6 préchauffé à 60°C sont ajoutés sur la résine suivi d'une incubation de 10 min à 60ºC. L'élution est récupérée par centrifugation de 10 sec à 10000 g à température ambiante. Cette étape est répétée quatre fois, afin de s'assurer que tous les ARNt retenus sur la sonde sont bien élués. Après précipitation à l'éthanol, les culots sont resuspendus dans un volume total d'eau de 15 µL. L'ARNt est quantifié et analysé sur un gel dénaturant (12% acrylamyde/bisacrylamyde 19 :1, 8 M urée, TBE 1 x) pour évaluer sa pureté.

*e. Analyse par spectrométrie de masse*

Philippe Wolf et Laura Antoine ont vérifié les séquences, la pureté, et la présence des modifications post-transcriptionnelles dans chaque ARNt purifié par spectrométrie de masse. Pour cela, chaque ARNt a été purifié sur gel dénaturant et digéré avec la RNase T1. Les fragments générés sont séparés sur une micro HPLC (NanoAquity, Waters) et analysés sur un

spectromètre (Synapt G2, Waters). Ainsi, la séquence et la présence des modifications post-transcriptionnelles de l'ARN peuvent être vérifiées.

### *3.2. Extraction d'ARN total de Plasmodium berghei*

Le sang de quatre souris ayant une parasitémie comprise entre 10 et 20% est nécessaire pour obtenir environ 8 µg d'ARN total de *P. berghei*, débarrassé des contaminants de l'hôte.

#### *a. Purification des globules rouges parasités*

Environ 1 mL de sang correspond à une ponction cardiaque réalisée à partir d'une souris impaludée. Le sang est centrifugé 10 min à 450 g pour séparer les érythrocytes du plasma sanguin (Figure 17A). Le culot de globules rouges est resuspendu dans 4,5 mL de PBS (9 volumes de culot) dans un Falcon de 15 mL. 9 mL de percoll 60% (60% percoll, PBS 1 X) sont délicatement déposés sous l'échantillon (Figure 57B1). Le tube est immédiatement centrifugé à 4ºC pendant 20 min à 1450 g dans un rotor oscillant (Figure 57B2). Les globules rouges infectés (GRi) ayant une moindre densité que les globules rouges non infectés, ceux-ci ne traversent pas le coussin de percoll. L'interphase contenant les GRi est récupérée et centrifugée à 4ºC pendant 10 min à 450g (Figure 57C1), et le culot est lavé avec 2 mL de PBS.

#### *b. Lyse globules rouges parasités*

Le culot de GRi est resuspendu dans 2 mL de tampon de lyse 1 x (10 mM $KHCO_3$, 150 mM $NH_4Cl$ et 0,1 mM EDTA) préalablement refroidit, et incubé 15 min dans la glace. Le tampon de lyse permet de lyser spécifiquement la membrane plasmique érythrocytaire et la vacuole parasitophore, pour ainsi libérer le parasite. Une centrifugation de 5 min à 100 g permet de culotter les débris cellulaires, alors que les parasites restent en suspension. Pour récupérer les parasites, le surnageant est centrifugé 5 min à 6000 g (Figure 17C2). Les parasites sont ensuite lavés trois fois dans 2 mL de PBS 1 x.

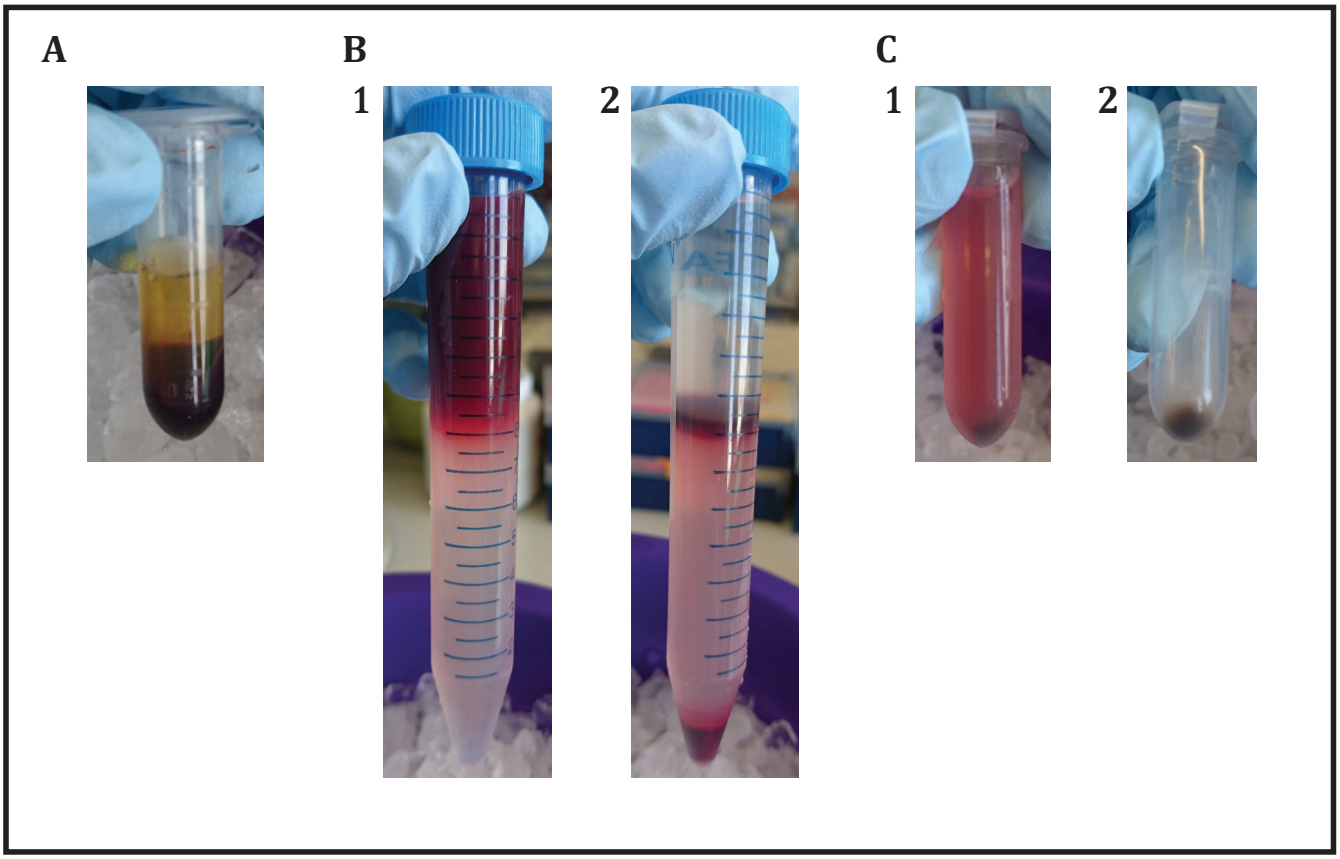



**Figure 17 : Les différentes étapes de la purification des stades sanguins de *P. berghei*.** **(A)** Séparation du plasma sanguin des cellules sanguines. **(B)** coussin de percoll 60% **(1)** avant et **(2)** après centrifugation. **(C)** Globules rouges **(1)** non lysés et **(2)** les parasites culottés.

*c. Elimination des ARN contaminants (de l'hôte)*

Un traitement à la RNase A est réalisé dans le but de dégrader les ARN de souris qui restent fixés sur la surface externe des parasites. Pour cela, le culot de parasites est resuspendu dans 1 mL de PBS 1 x en présence de 2,5 µg de RNase A. Après 15 min d'incubation à température ambiante, les parasites sont culotés (10 min à 2000 g et 4ºC), et immédiatement utilisés pour extraire les ARN totaux.

*d. Extraction de l'ARN*

Le culot de parasites est repris dans 200 µL de PBS 1x, 20 µL d'acide acétique 5 N et 750 µL de TriReagent. L'ARN est extrait selon le protocole Sigma. Le culot d'ARN final est resuspendu dans 10 à 25 µL d'eau et quantifié.

### 3.3. Production de transcrits d'ARNt in vitro

J'ai cloné les gènes correspondants aux transcrits d'ARNt$^{Ala}$, ARNt$^{Arg}$, ARNt$^{Leu}$, ARNt$^{Ser}$, ARNt$^{Thr}$ et ARNt$^{Tyr}$-1 humains Tableau 12). Les clones des ARNt$^{Asn}$, ARNt$^{Asp}$, ARNt$^{Lys}$ et ARNt$^{Tyr}$-2 étaient déjà disponibles dans l'équipe.

**Tableau 12 : Séquences des ARNt humains transcrits *in vitro*.**

| ARNt | Séquence | Ribo zyme | Enz yme |
|---|---|---|---|
| ARNt$^{Ala}$ | GGGGGTGTAGCTCAGTGGTAGAGCGCGTGCTTAGCATGCACGAGGCCCCGGGT TCAATCCCCGGCACCTCCACCA | non | Nsi1 |
| ARNt$^{Arg}$ | GGGCCAGTGGCGCAATGGAtAACGCGTCTGACTACGGATCAGAAGAtTCCAGG TTCGACTCCTGGCTGGCTCGCCA | non | Nsi1 |
| ARNt$^{Asn}$ | GTCTCTGTGGCGCAATCGGCTAGCGCGTTTGGCTGTTAACTAAAAGGTTGGTG GTTCGAACCCACCCAGAGGCG | oui | BstN1 |
| ARNt$^{Asp}$ | TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGT TCGATTCCCCGACGGGGAG | oui | BstN1 |
| ARNt$^{Leu}$ | GGTAGCGTGGCCGAGCGGTctAAGGCGCTGGATTTAGGCTCCAGTCTCTTCGG AGGCGTGGGTTCGAATCCCACCGCTGCCACCA | non | BstN1 |
| ARNt$^{Lys}$ | GGCCGGATAGCTCAGTCGGTAGAGCATCAGACTTTTAATCTGAGGGtCCAGGG TTCAAGTCCCTGTTCGGCCG | oui | Nsi1 |
| ARNt$^{Ser}$ | GTAGTCGTGGCCGAGTGGTTAAGGCGATGGACTAGAAATCCATTGGGGTTTCC CCGCGCAGGTTCGAATCCTGCCGACTACG | non | BstN1 |
| ARNt$^{Thr}$ | GGCGCGGTGGCCAAGTGGTAAGGCGTCGGTCTCGTAAACCGAAGAtCACGGGT TCGAACCCCGTCCGTGCCTCCA | non | BstN1 |
| ARNt$^{Tyr}$-2 | CCTTCGATAGCTCAGCTGGTAGAGCGGAGGACTGTAGATCCTTAGGTCGCTGG TTCAATTCCGGCTCGAAGGA | oui | BstN1 |
| ARNt$^{Tyr}$-1 | GGGGGCATAGCTCAGTGGTAGAGCATTTGACTGCAGATCAAGAGGtCCCTGGT TCAAATCCAGGTGCCCCCTCCA | non | Nsi1 |

Ce tableau précise si un ribozyme a été utilisé lors de la production, et quelle enzyme de restriction a été utilisée pour la linéarisation du plasmide d'ADN avant la transcription.

Pour les ARNt$^{Asn}$, ARNt$^{Asp}$, ARNt$^{Lys}$ et ARNt$^{Tyr}$-2, un ryboszyme a été introduit en 5' de la séquence de l'ARNt comme détaillé dans l'article Fechter, 1998. Cette astuce facilite la transcription des séquences 5' riches en C et T, et une fois transcrit, le rybozyme va s'auto-cliver et libérer un transcrit d'ARNt de la bonne taille.

*a. Conception des cassettes de clonage*

Les séquences d'ADN codant pour les ARNt alanine (trAla), arginine (trArg), leucine (trLeu), serine (trSer), thréonine (trThr) et tyrosine-1 (trTyr-1) ont étés produits par hybridation des oligonucléotides 1b à 10 (Figure 58) et ont été clonés dans le plasmide pUC119 entre les sites BamHI et HindIII, Chaque construction contient un site de restriction BstNI ou Nsi1 qui permet de produire une matrice de transcription finissant par la séquence GGT (transcrit en CCA 3' terminal).

```
trAla
        1b                  Ala-2                Ala-3                Ala-4                  Ala-5
AGCTTAATACGACTCACTATAGGGGGTGTAGCTCAGTGGTAGAGCGCGTGCTTAGCATGCACGAGGCCCCGGGTTCAATCCCCGGCACCTCCACCATGCATG
    ATTATGCTGAGTGATATCCCCCACATCGAGTCACCATCTCGCGCACGAATCGTACGTGCTCCGGGGCCCAAGTTAGGGGCCGTGGAGGTGGTACGTACCTAG
            Ala-6              Ala-7               Ala-8                 Ala-9                Ala-10

trArg
        1b                    Arg-2                  Arg-3                  Arg-4                  Arg-5
AGCTTAATACGACTCACTATAGGGCCAGTGGCGCAATGGATAACGCGTCTGACTACGGATCAGAAGATTCCAGGTTCGACTCCTGGCTGGCTCGCCATGCATG
    ATTATGCTGAGTGATATCCCGGTCACCGCGTTACCTATTGCGCAGACTGATGCCTAGTCTTCTAAGGTCCAAGCTGAGGACCGACCGAGCGGTACGTACCTAG
              Arg-6                  Arg-7                  Arg-8          Arg-9                  Arg-10

trLeu
        1b              Leu-2                  Leu-3                          Leu-4                  Leu-5
AGCTTAATACGACTCACTATAGGTAGCGTGGCCGAGCGGTCTAAGGCGCTGGATTTAGGCTCCAGTCTCTTCGGAGGCGTGGGTTCGAATCCCACCGCTGCCACCAG
    ATTATGCTGAGTGATATCCATCGCACCGGCTCGCCAGATTCCGCGACCTAAATCCGAGGTCAGAGAAGCCTCCGCACCCAAGCTTAGGGTGGCGACGGTGGTCCTAG
         Leu-6                        Leu-7          Leu-8                  Leu-9                  Leu-10

trSer
       1b                     Ser-2        Ser-3                      Ser-4                      Ser-5
AGCTTAATACGACTCACTATAGTAGTCGTGGCCGAGTGGTTAAGGCGATGGACTAGAAATCCATTGGGGTTTCCCCGCGCAGGTTCGAATCCTGCCGACTACGCCAG
    ATTATGCTGAGTGATATCATCAGCACCGGCTCACCAATTCCGCTACCTGATCTTTAGGTAACCCCAAAGGGCGCGTCCAAGCTTAGGACGGCTGATGCGGTCCTAG
              Ser-6                Ser-7                      Ser-8                  Ser-9          Ser-10

trThr
       1b                    Thr-2                  Thr-3          Thr-4                  Thr-5
AGCTTAATACGACTCACTATAGGCGCGGTGGCCAAGTGGTAAGGCGTCGGTCTCGTAAACCGAAGATCACGGGTTCGAACCCCGTCCGTGCCTCCAG
    ATTATGCTGAGTGATATCCGCGCCACCGGTTCACCATTCCGCAGCCAGAGCATTTGGCTTCTAGTGCCCAAGCTTGGGGCAGGCACGGAGGTCCTAG
            Thr-6                  Thr-7          Thr-8                  Thr-9          Thr-10

trTyr-1
       1b              Tyr1-2                 Tyr1-3                 Tyr1-4                 Tyr1-5
AGCTTAATACGACTCACTATAGGGGGCATAGCTCAGTGGTAGAGCATTTGACTGCAGATCAAGAGGTCCCTGGTTCAAATCCAGGTGCCCCCTCCATGCATG
    ATTATGCTGAGTGATATCCCCCGTATCGAGTCACCATCTCGTAAACTGACGTCTAGTTCTCCAGGGACCAAGTTTAGGTCCACGGGGGAGGTACGTACCTAG
              Tyr1-6         Tyr1-7                 Tyr1-8                 Tyr1-9         Tyr1-10
```

**Figure 58 : Cassettes utilisés pour la synthèse des transcrits alanine, arginine, leucine, serine, thréonine et tysorine-1.**

*b. Transcription*

Le plasmide est linéarisé avec l'enzyme Bstn1 ou Nsi1 (Tableau 12), et le gène d'intérêt est transcrit. 50 µg d'ADN linéaire sont incubés 3 h à 37ºC en présence de 4 mM de chaque nucléotide, 8 mM GMP, 40 mM Tris-HCl pH 8,1 (à 37°C), 30 mM $MgCl_2$, 0,01% Triton, 1 mM Spermidine, 25 mM DTE, et 10 µL de $T_7$ RNA polymérase. A la fin de la transcription, Les transcrits possédant un ribozyme (Fechter, *et al.*1998), sont dilués dans 4 volumes de tampon de clivage (40 mM $MgCl_2$ et 30 mM Tris pH 8,1) et incubés 1 h à 60ºC.

Les transcrits ont étés purifiés sur un grand gel dénaturant (33 cm x 39 cm x 0,5 cm). La migration est effectuée sur la nuit, à 600 v. La bande de gel contenant le transcrit est découpée, et le transcrit est électroélué dans 800 µL de TBE 1 X pendant 2 fois 1 h à 150 v, et précipité.

### *3.4. Renaturation*

Les ARNt produits *in vivo*, comme les transcrits *in vitro* sont systématiquement renaturés avant chaque expérience : Ils sont placés 2 min à 65ºC, suivi d'un refroidissement à température ambiante pendant 10 min.

## 4. Etudes structurales et fonctionnelles des ARN

### *4.1. Marquages radioactifs des acides nucléiques*

*a. Marquage en 5'*

Au cours de mes expériences, j'ai été amenée à marquer en 5' deux sortes d'acides nucleiques : des ARNt qui sont utilisés dans les études structurales (voir chapitre 0.) et des desoxyoligonucleotides utilisés comme sondes dans des expériences de Northern blot (voir chapitre 0.).

Les ARNt natifs possèdent un phosphate en 5'. La première étape consiste donc à enlever ce phosphate à l'aide de l'enzyme FastAP. 80 pmol d'ARNt sont diluées dans un volume total de 20 µL contenant 2 µL de tampon FastAP 10 x et 1 µL d'enzyme (1 U/µL), et incubés à 37ºC pendant 10 min. La réaction est arrêtée par une extraction phénolique et les ARN déphosphorylés sont précipités. Cette étape de déphosphorilation n'est pas nécessaire dans le cas des sondes d'ADN.

Le marquage est réalisé dans un volume total de 10 µl, en présence de 30 pmol de l'acide nucléique, du tampon PNK-A, de 10 U/µL de $T_4$PNK et de 30 µCi de [$\gamma^{32}$P]-ATP. Le tout est incubé 20 min à 37ºC. Les ARN sont purifiés sur gel dénaturant (12% acrylamyde/bisacrylamide 19 :1, urée 8 M, TBE 1 x) pendant 2 h à 150 v. La bande contenant l'ARN radioactif est révélée par autoradiographie et découpée. L'ARN est élué sur la nuit à 4ºC sous agitation dans 400 µL de tampon Maxam-Gilbert (0,2 M acétate d'ammonium, 10 mM acétate de magnésium, 20 mM Tris-HCl pH 7,4, 0,1% SDS), et précipité sous éthanol. Les sondes d'ADN sont purifiées sur Microspin G-25 (GE Healthcare).

La radioactivité est quantifiée par comptage Cerenkov et les marquages sont conservés à -20°C jusqu'à leur utilisation.

*b. Marquage en 3' des ARNt*

Le marquage des ARNt en 3' se fait en deux étapes. Pendant la première étape, le CCA en 3' est enlevé partiellement : 100 pmol d'ARNt sont incubés 10 min à 20ºC dans un milieu composé de 180 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 0,5 µg de phosphodiestérase de venin de cobra. La réaction est arrêtée par une extraction phénolique, et l 'ARNt est précipité à l'éthanol.

La deuxième étape consiste à reconstruire le CCA en 3' en présence de [$\alpha^{32}$P]-ATP. L'ARN est incubé avec 3,3 µg de CCAse d'*E. coli* pendant 30 min à 37ºC dans 50 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 45 µM CTP, 7,5 mM DTE, 50 µCi [$\alpha^{32}$P]-ATP. L'ARNt marqué est purifié sur une colonne NAP-5 (GE Healthcare), et précipité à l'éthanol.

La radioactivité est quantifiée par comptage Cerenkov et les marquages sont conservés à -20°C jusqu'à leur utilisation.

***4.2. Etudes du Repliement en solution des ARNt***

Deux types de sondes ont été utilisées : (i) l'acétate de plomb (plomb ou Pb) et (ii) des sondes enzymatiques telles que les RNases T1, V1 et S1. Toutes ces expériences ont été réalisées avec 50000 cpm d' ARNt marqué en 5'.

a. *Coupure au plomb*

L'ion plomb catalyse l'hydrolyse du squelette sucre-phosphate de l'ARNt uniquement si l'hydroxyle situé en 2' dans la ribose est accessible (Sundaralingam, *et al.*, 1984). La coupure ne dépend donc que de la structure tridimensionnelle de l'ARNt. Un clivage unique dans chaque molécule d'ARNt marquée radioactivement, suivi de la migration des fragments sur un gel, permet la visualisation des zones accessibles au plomb, et donc permet l'étude de du repliement de la molécule.

L'ARN radioactif est incubé 6 min à température ambiante avec 2 µM d'ARN froid dans un milieu contenant 50 mM HEPES-NaOH pH 7,5, 5 mM acétate de magnésium, 50 nM acétate de potassium et différentes concentrations de plomb (concentrations finales : 2,5 mM, 5 mM et 10 mM). La réaction est arrêtée en ajoutant 16 µL de tampon STOP (30 mM EDTA, 62,5 µg/mL glycogène, 62,5 ng/µL ARN total de E. coli, 375 mM acétate de sodium). Contrairement aux autres composants, le glycogène ne participe pas à l'arrêt de la réaction mais est utile pour la précipitation des fragments d'ARN. Après précipitation à l'éthanol, les

échantillons sont déposés sur un gel dénaturant 12% acrylamyde/bisacrylamyde 19:1, 8 M urée, TBE 1 x (33 cm x 39 cm x 1 mm), qui migre entre 2 et 3 h à 1600 v.

*b. Utilisation de sondes enzymatiques*

Les enzymes utilisées pour cette analyse sont des RNases qui ont chacune leur spécificité de coupure. La RNAse $S_1$ nécessite du Zn comme cofacteur (1 mM $ZnCl_2$,) et coupe les régions simple brin des ARN, comme la RNase $T_1$ alors que la RNAse $V_1$ reconnait spécifiquement les régions appariées ou fortement structurées. 50000 cpm d'ARN sont incubés dans un volume total de 10 µl en présence de 1 µM du même ARN non marqué, 25 mM Tris-HCl pH 7,5, 10 mM $MgCl_2$, et 50 mM NaCl, et deux quantités d'enzymes (0,13 et 0,26 U RNase $T_1$, 5,4 et 7,6 U RNase $S_1$, 0,7 et 1,4 mU RNase $V_1$) pendant 10 min à 25ºC. La réaction est arrêtée en ajoutant 21 µL de tampon STOP composé de 30 mM EDTA, 62,5 µg/mL glycogène, 62,5 ng/µL ARN total de E. coli, 375 mM acétate de sodium, puis les ARN sont précipités à l'éthanol. Les échantillons sont ensuite analysés sur gel dénaturant comme indiqué dans le paragraphe précédent.

Pour assigner les bandes obtenues, chaque expérience inclue une hydrolyse alcaline (un échelle qui permet de repérer chaque nucléotide de la séquence), et une digestion à la RNAse T1 en conditions dénaturantes (permet de déterminer la position de tous les G présents dans la séquence).

L'hydrolyse alcaline est réalisée avec 50000 cpm d'ARN marqué en présence de 2 µM du même ARN non marqué dans 40 mM $NaHCO_3$. La réaction est incubée 5 min à 80ºC et est arrêtée en rajoutant 1 µg de glycogène et 25 µL d'éthanol pour précipiter l'ARN.

L'hydrolyse avec l'enzyme $T_1$ en conditions dénaturantes est réalisée avec 50000 cpm d'ARN, 1 µM du même ARN froid, et du tampon spécifique pour l'enzyme (10 mM citrate de sodium pH 4,5, 3,2 M urée, 0,4 mM EDTA). La réaction est préchaufée à 60ºC pendant trois minutes et 2 U de $T_1$ sont rajoutées pour 1 min et 15 sec à 60ºC. L'ARN est précipité à l'éthanol en présence de 1 µg de glycogène.

***4.3. Etudes fonctionnelles***

*a. Northern blots*

L'identification des ARNt extraits des stades sanguins du parasite *P. berghei* (voir *0.*) a été faite par la méthode de Northern blot, en utilisant différentes sondes conçues pour faire

la différence entre les ARN provenant de *P. berghei* et ceux provenant de *Mus musculus* (les séquences des sondes sont indiquées dans la partie Résultats). Dans un premier temps, les différents ARN contenus dans 1 ou 4 µg d'ARN total sont séparés sur gel dénaturant 12% acrylamyde/bisacrylamyde (19 :1), urée 8 M, TBE 1 x. Les ARN sont transférés sur membrane de nylon chargée positivement (BrightStar™ Plus, Ambion). Pour cela, la membrane et le gel sont mis entre deux feuilles de papier Whatman imbibées de TBE 1x stéril, et un courant de 200 mA est appliqué pendant 30 min. Les ARN sont ensuite crosslinkés dans un four (Stratagène) pendant 2 min à 120 mJ. La membrane a été incubée pendant 2 h à 60ºC, en présence du tampon d'hybridation (UltraSensitive Buffer, Ambion).

La membrane est ensuite hybridée sur la nuit à 60ºC avec 5 mL de tampon d'hybridation contenant 3300000 cpm de sonde radioactive, et 0,25 mg d'ADN de sperme de saumon pour bloquer les interactions aspécifiques. Le lendemain, le tampon d'hybridation est retiré, et la membrane est rincée avec 5 mL de SSC 2X (300 mM chlorure de sodium et 34 mM de citrate de sodium, 0,1% SDS) et lavée deux fois pendant 20 min à 60ºC avec 5 mL du même tampon.

La membrane est thermoscellée entre deux feuilles de plastique et est mise à exposée avec un film sensible (Carestream) pendant au moins 5 h.

*b. Interactions ARN/protéines*

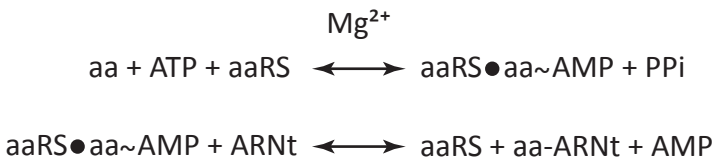
Les interactions ARN/protéine ont été mises en évidence et leurs affinités ont été quantifiées par la méthode de retard sur gel. Le complexe ARN/protéine est formé en incubant l'ARN (10000 cpm d'ARN marqué en 3') avec différentes concentrations de protéine dans un volume total de 20 µl pendant 20 minutes à 4ºC dans un milieu réactionnel composé de 25 m M d'HEPES-NaOH pH 7,5, 75 mM de KCl, 5 mM de $MgCl_2$, 20 nM d'oligo dT et 10% de glycérol. Le milieu réactionnel est ensuite déposé sur un gel natif (7% acrylamyde/bisacrylamyde 37,5 :1, TBE 1 x). La migration est faite pendant 3h30 min à 140 v à 4ºC

Une variante de l'expérience de retard sur gel consiste à faire des compétitions entre deux ARN, le premier est marqué et présent à concentration constante et limitante, le second, appelé compétiteur, est « froid » et est utilisé à différentes concentrations. Pour être dans des conditions de compétitions, l'ARN marqué est incubé avec une concentration de protéine définie pour obtenir au maximum la moitié de l'ARN dans le complexe. Dans le cas de la compétition entre l'ARNt total de cellules HeLa et le transcrit trAla, 100 nM de *Pf*tRip, 20 nM

d'ARNt total (dont 10000 cpm) et différentes concentrations de trAla (7,5 nM, 15 nM, 30 nM, 60 nM, 120 nM et 250 nM) sont incubés pendant 20 min à 4ºC dans le milieu réactionnel utilisé pour l'expérience de retard sur gel. Les échantillons sont ensuite déposés sur un gel natif comme indiqué dans le paragraphe ci-dessus

### *4.4. Aminoacylation des ARNt*

La réaction d'aminoacylation se fait en deux étapes. La première étape correspond à l'activation de l'acide aminé et la deuxième étape consiste au transfert de l'acide aminé activé sur l'ARNt. Chacune de ces étapes peut être mesurée *in vitro*.




*a. Activation de l'acide aminé*

Cette étape est mesurée par la réaction d'échange ATP/PPi qui permet de quantifier la réaction inverse, c'est à dire la formation d'ATP radioactif à partir de PPi radioactif et d'aminoacyl-adenylate. La *Pfapi*TyrRS sauvage ou mutée (2 µg) sont incubées avec 2 mM de [$^{32}$P]-PPi (2 cpm/pmol) pendant 2 minutes à 37ºC dans un volume final de 200 µL contenant 100 mM HEPES pH 7,5, 10 mM $MgCl_2$, 2 mM ATP, 4 mM d'inhibiteur de la pyrophosphatase (KF), et 5 mM de L-tyrosine. Aux temps 4, 8, 12 et 16 min, la réaction est interrompue en mélangeant une aliquote de 50 µL avec 200 µL de charbon actif (2% charbon, 15% acide perchlorique, 200 mM PPi). L'ATP contenu dans le milieu va être adsorbé par le charbon actif. Le mélange est dilué dans 2 mL d'eau et déposé sur un filtre de 25 mm de diamètre (GF/C, Whatman). Le charbon retenu sur le filtre est ensuite lavé trois fois avec 5 mL d'eau déminéralisé et avec 5 mL d'éthanol. Le filtre est séché à l'air, et la radioactivité du [$^{32}$P]-ATP retenu sur le filtre est quantifiée en présence de 2,5 mL de liquide de scintillation.

*b. Aminoacylation des ARNt*

Pour déterminer l'efficacité du transfert de l'acide aminé sur l'ARNt$^{Tyr}$ apicoplastique et ses mutants par la *Pfapi*TyrRS, la réaction est réalisé en présence d'acide aminé tyrosine marqué radioactivement avec du $^{14}$C (L-[$^{14}$C]-tyrosine). Cet acide aminé (10 µM à

486 mCi/mmol) est incubé 2 min à 37ºC avec plusieurs concentrations d'ARNt, dans un milieu qui favorise la réaction (50 mM HEPES-KOH pH 7,6, 25 mM KCl, 12 mM $MgCl_2$, 2,5 mM ATP, 0,2 mg/mL sérum d'albumine bovine, 1 mM spermine), comme indiqué dans l'article Cela, 2018. La réaction d'aminoacylation commence au moment où *Pfapi*TyrRS est ajoutée au milieu réactionnel. Aux différents temps donnés, 15 µL sont prélevés et déposés sur une étiquette de papier whatman qui est ensuite lavé pendant 10 min dans du TCA 5% froid pour faire précipiter l'ARNt sur le papier. Pour ce premier lavage, le TCA est saturé en tyrosine, ce qui diminue la fixation non spécifique de l'acide aminé à l'étiquette. Après deux lavages dans du TCA 5% (10 min dans la glace), les étiquettes sont déshydratées dans de l'éthanol et séchées. La radioactivitée retenue par les étiquettes est quantifiée en présence de 2,5 mL de liquide de scintillation.

## 5. Techniques de sélection *in vitro*

### *5.1. MIST*

*a. Le principe*

La technique du MIST (Eriani, *et al.*, 2015) a été utilisée en collaboration avec son inventeur le Dr. Renaud Geslain (College of Charleston, Caroline du Sud, USA), pour identifier les ARNt en fonction de leur affinité pour *Pf*tRip. Cette technique consiste dans un premier temps à isoler le pool d'ARNt fixé sur *Pf*tRip à l'aide d'une expérience de retard sur gel, puis dans un deuxième temps à identifier et quantifier ces ARNt à l'aide d'une puce conçue spécialement pour hybrider les ARNt (tRNA microarray ou puce à ARNt).

**Tableau 13 : Séquences des sondes conçues pour hybrider les ARNt cytosoliques humains.**

| Sonde | ARNt | Séquence |
|---|---|---|
| Ala-1h | ala(IGC) | TGGAGAATGYGGGCGTCGATCCCRCTACCTCTYGCATGCTAAGCRAGCGCTCTACCRCCTGAGCTAATTCCCC |
| Ala-2h | ala(hGC) | TGGAGRTGCCGGGGATYGAACCCGGGRCCTCRTRCATGCSAAGCAYGCGCTCTACCACTGAGCTACAYCCCC |
| Ala-3h | ala(TGC) | TGGAGAAGCAGGGAATTGAACCCTGTACTTAAGTTTGCAAAACTTATGTTTAACCATGTAAACTACATCCCC |
| Arg-1h | arg(ICG) | CGAGCCAGCCAGGAGTCGAACCTRGAATCTTCTGATCCGTAGTCAGACGCGTTATCCATTGCGCCACTGGCCC |
| Arg-2h | arg(yCG) | CAACCACGRAGGGACTCGAACCCTCAATCTTCTGATCCGRARTCAGACGCCTTATCCATTAGGCCACGCGGCC |
| Arg-3h | arg(CCT) | TACCCCAGGTGGGACTCGAACCCACAATCCCTGGCTTAGGAGGCCAGTGCCTTATCCATTAGGCCACTGGGGC |
| Arg-4h | arg(TCT) | CGACTCTGCYGGGACTCGAACCCGWAACCTTTGAATTAGAAGTCCAATGCGCTATCCATTGCGCCACAGAGCC |
| Asn-1h | asn(GTT) | CGTCCCTGGGTGGGCTCGAACCACCAACCTTTCGGTTAACAGCCGAACGCGCTAACCGATTGCGCCACAGAGAC |
| Asp-1h | asp(GTC) | CTCCCCGTCGGGGAATCGAACCCCGGTCTCCCGCGTGACAGGCGGGGATACTCACCACTATACTAACGAGGA |
| Cys-1h | cys(GCA) | AGGGGGCACCYGGATTTGAACCRGGGACCTCTTGATCTGCAGTCAAATGCTCTACCMCTGAGCTATACCCCC |

| | | |
|---|---|---|
| Gln-1h | gln(yTG) | AGGGGGCACCYGGATTTGAACCRGGGACCTCTTGATCTGCAGTCAAATGCTCTACCMCTGAGCTATACCCCC |
| Glu-1h | glu(yTC) | TTCCCTGACCGGGAATCGAACCCGGGCCWMGGCGGTGARAGCGCCGAATCCTARCCACTAGACCACCAGGGR |
| Glu-2h | glu(TTC) | TTCCCAYACCGGGAGTCGAACCCGGGCCRCCTGGGTGAAAACCAGGAATCCTAACCGCTAGACCATRTGGGA |
| Glu-3h | gly(TCC2) | TTCGTTACCCGGGAATCGAACCCGCGTTAACTGCTTGAAAGGCAGCTGTGCTCACCACTACACCACCAACGC |
| Gly-1h | gly(sCC) | TGCATTGGCCRGGAATYGAACCCGGGYCTCCCRCGTGGWAGGCGAGAATTCTACCACTGMACCACCMAYGC |
| Gly-2h | gly(TCC) | TGCGTTGGCCGGGAATCGAACCCGGGTCAACTGCTTGGAAGGCAGCTATGCTCACCACTATACCACCAACGC |
| His-1h | his(GTG) | TGCCGTGACTCGGATTCGAACCGAGGTTGCTGCGGCCACAACGCAGAGTACTAACCACTATACGATCACGGC |
| Ile-1h | ile(rAT) | TGGCCMGTACGGGGATCGAACCCGCGACCTTGGCGTTATTAGCACCACGCTCTAACCAACTGAGCTAACCRGCC |
| Ile-2h | ile(TAT) | TGCTCCAGGTGAGGMTCGAACTCACAACCTCGGCATTATAAGTACCGCGCGCTAACCGATTGCGCCACTGGAGC |
| Leu-1h | leu(wAG) | TGGCAGYGGTGGGATTCRAACCCACGCCYCCGAARAGACTGGAGCCTTAATCCAGCGCCTTAGACCRCTCGGCCACRCTA |
| Leu-2hn | leu(CAG) | TGTCARAAGTGGGATTCGAACCCACGCCTCCAGWGGAGGWTGCGATTTGAACGCAGCGCCTTAAGACYGCTCGGCCATCC |
| Leu-3h | leu(CAA) | TGTCAGAAGTGGGATTCGAACCCACGCCTCCATTGGAGACCAGAACTTGAGTCTGGCGCCTTAGACCACTCGGCCATCC |
| Leu-4h | leu(TAA1) | TACCRGGAGTGGGGYTCGAACCCACGCGGACAYCSGTCCATTGGATCTTAAGTCCAACGCCTTAACCACTCRGCCATCC |
| Leu-5h | leu(CAA) | TGTTCCAAGACGGATTCGAACCGTCACTAATACAAGGTTTGAGCTTGCATCCTCTGCCAATTGGGATACTGGAAC |
| Leu-6h | leu(TAA2) | TGTTAAKRAGAGGAGTTGAACCTCTGATTATAAARTTTTAAGTYTTATGCAATTRCCRGGCTCTGCCATCTTAAC |
| Lys-1h | lys(CTT) | CGCCCAACGTGGGGCTCGAACCCACGACCCTGAGATTAAGAGTCTCATGCTCTACCGACTGAGCTAGCCGGGC |
| Lys-2h | lys(TTT1) | CGCCYGAACAGGGACTTGAACCCTGGACCCTCAGATTAAAAGTCTGATGCTCTACCRACTGAGCTATCCRGGC |
| Lys-3h | lys(TTT2) | CACCCAAAAAGGGACATGAACCTTGGACCCTCAGATTAAAAGTCTGATGCTCTACTGACTGAGCTATCTGGGT |
| Lys4H | lys(TTT3) | CGCCTGGACAGGGACTTGAACCCTAGACCCTCAGATTAAAAGTCTGATGTTCTACCTACTGAGCTACCCAGGT |
| mAla2H | ala(hGC2) | TGGAGGTGTCGGGGATCGAACCGAGGCCTCATACATGCAAAGCATGCGCTCTACCACTGAGCTACACCCCC |
| mArg2H | arg(yCG2) | AACCACGAAGGGATTCGAACCCTCAATCTTCTGATCCGAAGTCAGACGCCTTATCCATTAGGCCACGTGGTC |
| Met-1h | meti | TAGCAGAGGATGGTTTCGATCCATCGACCTCTGGGTTATGGGCCCAGCACGCTTCCGCTGCGCCACTCTGCT |
| Met-2h | mete | TGCCCYCTCTGAGGCTCGAACTCAGGACCTTCAGATTATGAGACTGACGCGCTGCCHRCTGCGCTAAGRRGGC |
| mMet2M | tRNA meti | TAGCAGAGGATGGTTTCGATCCATCGACCTCTGGGTTATGGGCCCAGCACGCTTCCGCTGCGCCACTCTGCT |
| Phe-1h | phe(GAA) | TGCYGAAACCCGGGATYGAACCAGGGACCTTTAGATCTTCAGTCTAACGCTCTCCCAACTGAGCTATTTCGGC |
| Pro-1h | pro(hGG) | GGGCTCGTCCGGGATTTGAACCCGGGACCTCTCGCACCCDAAGCGAGAATCATACCCCTAGACCAACGAGCC |
| Sec-1h | sel-Cys | TGACCACAAAGGGACTCAAACCCTCAATCTTCTGATCTGAAGTCAGACACCTTATCCATTAAGGCCACACGGTC |
| Sec-2hn | sel-Cys2 | AAGGTGGAATTGAACCACTCTGTCRCTAGACAGCTACAGGTTTGAAGCCTGCACCCCAGACCACTGAGGATCATCCGRGC |
| Ser-1h | ser(CGA) | GCTGTGAGCAGGATTTGAACCTGCGCGGGGARACCCCATTGGATTTCGAGTCCAACGCCTTAACCACTCGGCCATCACAG |
| Ser-2h | ser(wGA) | GTAGYCGGCAGGATTCGAACCTGCGCGGGGARACCCCAATGGATTTCSAGTCCATCGCCTTAACCACTCGGCCACGACTA |
| Ser-3h | ser(GCT) | GACGAGGRTGGGATTCGAACCCACGYGTGCAGAGCACAATGGATTAGCAGTCCATCGCCTTAACCACTCGGCCACCTCGT |
| Ser-4h | ser/cys | AGGGGGCACCTGGATTTGAACCAGGGACCTCTTGATCTCTAGTCAAATTCTCTACCCCTGAGCCATACACCC |
| Thr-1h | thr(mGT) | AGGCCCCRCTGGGAKTCGAACCCAGGATCTCCTGTTTACKAGACAGGCGCTTTAACCARCTAAGCCAYRGNGCC |
| Thr-2h | thr(TGT1) | AGGCCCCAGCGAGATTYGAACTCGCGACCCCTGGTTTACAAGACCAGTGCTCTAACCMCTGAGCTATGGAGCC |
| Thr-3h | thr(CGT1) | AGGCCCTGGTGAGATTTGAACTCACGACCCCTAGGTTTACGAGACCAGCGCTCCAACCGCTGAGCTACAGGGCC |
| Thr-4h | thr(TGT2) | TGCTGCTCACAGGACTCGAACCTGCATTTCCATCCTTACAAGGGATGTGTAATAGCCAATTATACGAAAGCAGC |
| Thr-5h | thr(CGT2) | AGGCACGGACGGGGTTCGAACCCGYGATCTTCGGTTTACGAGACCGACGCCTTACCACTTGGCCACCGCGCC |
| Thr6h | thr/meti | TAGCCAAGGATGGCTTCGATCCATTGACCTCTGGGTTACGGGCCCAGCACGCTTCCGCTGCACCACTCTGCC |
| Trp-1h | trp(CCA) | TGACCCCGACGTGATTTGAACACGCAACCTTCTGATCTGGAGTCAGACGCGCTACCGTTGCGCCACGAGGTC |
| Tyr-1h | tyr/cys | AGGGAGAACCTGGATTTGAACCAGGGACCTCTTGCTCTACAGTCAAAAAGCTCTAGCCCTGAGCTATACCCCC |
| Tyr-2h | tyr(GTA) | TCCTTCGAGCCGGASTCGAACCAGCGACCTAAGGATCTACAGTCCTCCGCTCTACCARCTGAGCTATCGAAGG |
| Val-1h | val(mAC) | TGTTTCYGCCYGGTTTCGAACCRGGGACCTTTCGCGTGTKAGGCGAACGTGATAACCACTACACTACRGAAAC |
| Val3H | val/ala | TGGAAGTGCTGGGGATCGAACCCAGAGCCTCATGAATGTTAAGCATACGCTCTACCACTGAGCTACACCCCC |
| Val-4h | val(TAC) | TGGTTCCACTGGGGCTCGAACCCAGGACCTTCTGCGTGTAAAGCAGACGTGATAACCACTACACTATGGAACC |

*b. Conception de la puce à ARNt*

48 sondes spécifiques à chaque famille d'ARNt humain (Tableau 13) ont été conçues et produites par le Dr. Geslain selon les protocoles indiqués dans l'article Emetu, *et al.*, (2018). Chacune a été disposée huit fois sur chaque puce (Figure 59), donc chaque expérience d'hybridation correspond à 8 répliques techniques.

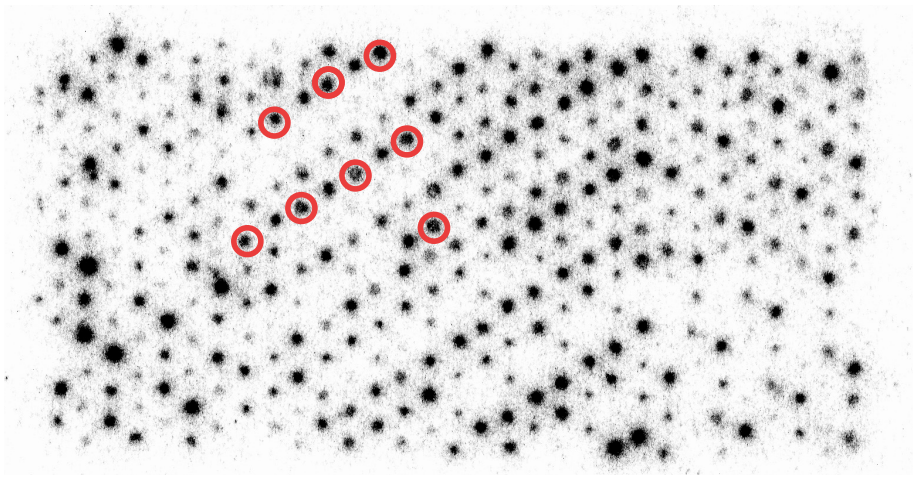


**Figure 59 : Une puce à ARNt hybridée à de l'ARNt total (purifié à partir de cellules HeLa).** Résultat de l'exposition d'une plaque d'imagerie (Fujifilm) pendant 90 min. Chaque tache noire représente la localisation d'une sonde. Entourées en rouge, la disposition d'une même sonde à huit emplacements différents sur la puce.

*c. Préparation des échantillons d'ARNt compléxés ou pas à tRip*

Nous avons pris le parti d'incuber *Pf*tRIp avec plusieurs concentrations d'ARNt total de cellules HeLa, toujours dans des conditions où l'ARNt est en excès par rapport à la protéine. 4 conditions ont été utilisées : dans un volume total de 100 µl contenant 25 m M d'HEPES-NaOH pH 7,5, 75 mM de KCl, 5 mM de $MgCl_2$, 20 nM d'oligo dT et 10% de glycérol, 4 µM *de Pf*tRip est incubée en présence de 2, 8, 20, et 40 µM d'ARNt total pendant 20 min à 4ºC. Les ARNt complexés ont été séparés des ARNt libres sur un gel natif (7% acrylamyde/bisacrylamyde (37,5 : 1), 1 x TBE) au cours d'une migration longue 4 h à 140 v. Les bandes de gels contenant l'ARNt complexé ou l'ARNt libre sont découpées et les ARNt sont élués selon la méthode Crush&Soak (300 mM acétate de sodium, 1 mM EDTA) en présence de 0,1% SDS (Petrov, *et al.*, 2013).

**Tableau14 : Conditions d'hybridation des ARNt à la puce et composition des tampons utilisés.**

| | Temps | Température | Tampons | |
|---|---|---|---|---|
| Conditionner les joints | 2 min | 75ºC | - | |
| Injection des ARNt marqués | | | | |
| Dénaturation | 5 min | 90ºC | 2 x SSC 0,01%SDS | |
| Hybridation | 3 h | 60ºC | 2 x SSC 0,01%SDS | |
| Lavage 1 | 30 sec | 50ºC | 2 x SSC 0,1% SDS | 2 |
| Lavage 2 | 30 sec | 42ºC | 0,1 x SSC 0,1% SDS | 2 |
| Lavage 3 | 30 sec | 42ºC | 0,1x SSC | 2 |

Composition du tampon SSC : 300 mM chlorure de sodium, 34 mM de citrate de sodium.

En parallèle, 2 expériences contrôles ont été effectuées en l'absence de protéine. Cette expérience permet de quantifier chaque ARNt dans le pool de départ.

Les ARNt ainsi purifiés sont marqués en 3' comme indiqué dans le paragraphe 0, et hybridés à la puce dans les conditions indiquées dans le Tableau 14. La composition des tampons utilisés est décrite dans l'article Emetu, *et al.* (2018).

*d. L'analyse des donnés*

Les puces radioactives sont exposées avec une Imaging plate (Fujifilm), qui est scannée (Typhoon FLA 7000, GE Healthcare) et l'intensité de la radioactivité est quantifiée avec ImageJ. Les résultats quantitatifs de chaque expériences effectuées en présence de *Pf*tRip ont été systématiquement comparés à une expérience contrôle (ARNt total ou la somme des ARNt liés et libres) pour déterminer l'enrichissement ou l'appauvrissement de certains ARNt soit dans le complexe soit dans l'ARNt qui ne se lie pas à tRip.

***5.2. SELEX***

*a. Le principe*

La technique du SELEX (Stoltenburg, *et al.*, 2007) consiste à sélectionner dans une bibliothèque d'aptamères, les molécules d'acides nucléiques susceptibles de se lier

spécifiquement à une protéine cible. Grâce à cette approche, j'ai recherché des aptamères d'ARN capables de mimer des ARNt en se fixant spécifiquement à *Pf*tRip (Figure 60).

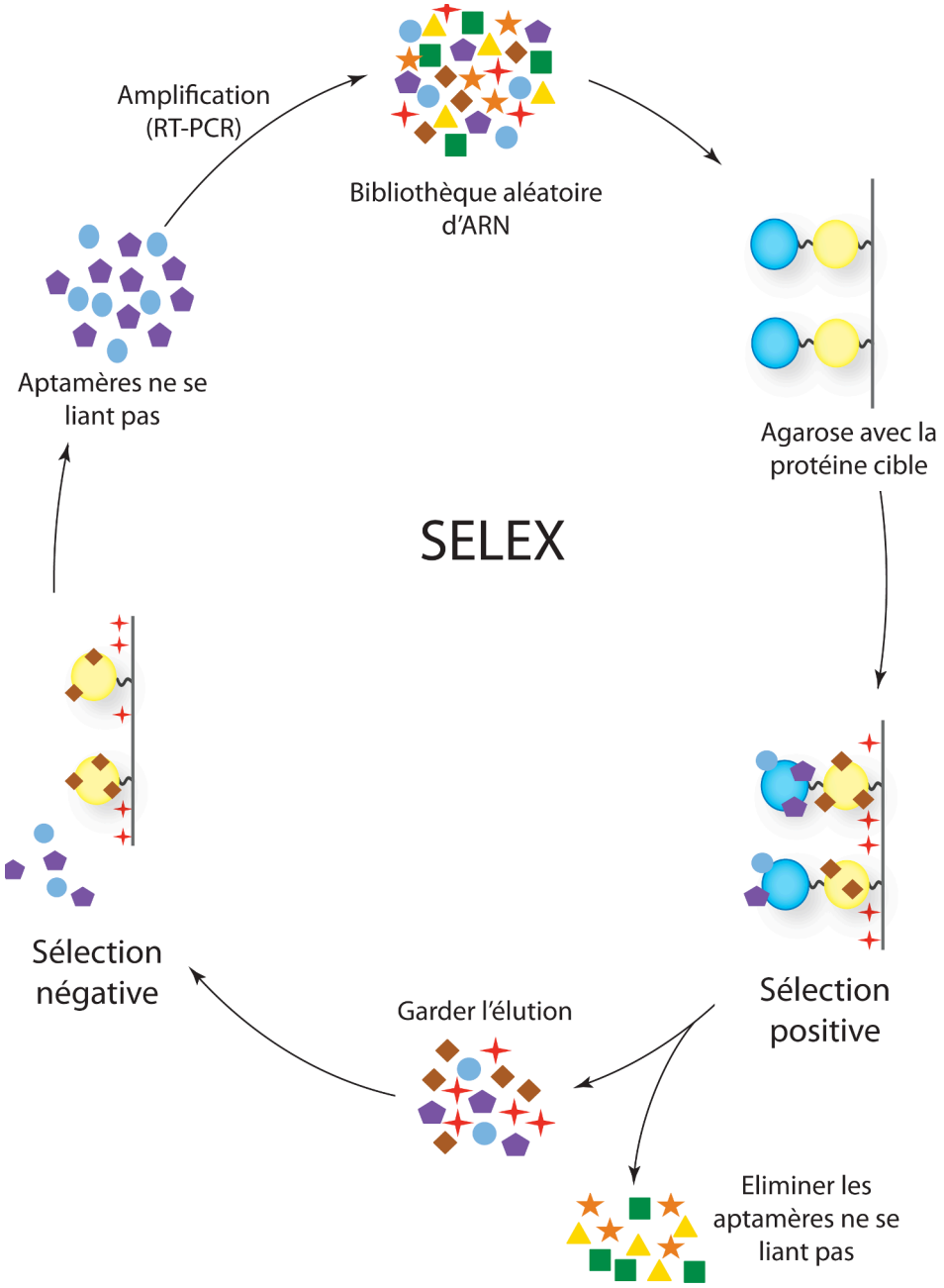



**Figure 60 : Pricipe du SELEX.** Les pentagones violets, cercles bleus, croix rouges, carrés verts, triangles jaunes, losanges marrons et étoiles oranges représentes différents aptamères. Les sphères jaunes (partie GST qui se fixe à la résine) et bleues (domaine C-terminal de *Pf*tRip) représentent les deux domaines de la protéine de fusion cible.

*b. Construction de la bibliothèque d'aptamères*

La bibliothèque d'ADN a été construite par PCR à partir des oligonucléotides décrits dans le Tableau 15. La séquence de la matrice contient en 5' le promoteur de la T7 ARN polymérase suivi de 25 nucléotides aléatoires et en 3' une séquence de 23 nucléotides utilisés

pour l'amplification par PCR. La synthèse de la bibliothèque ADN a été effectuée en utilisant 2 ng d'oligo matrice, 3 ng de chaque amorce dans un milieu contenant 75 mM Tris pH 8,8, 20 mM sulfate d'ammonium, 0,01% Tween 20, 2 mM $MgCl_2$, 60 µg de polymérase Taq, et 1 mM de chaque dNTP. Après 2 cycles de 1 min à 95ºC, 1 min à 58ºC, 10 min à 72ºC, le produit de la PCR a été purifié sur un gel 12% acrylamyde/bisacrylamyde (37,5 : 1), 8 M urée, 1 x TBE (migration pendant la nuit à 200 v et 50 mA).

**Tableau 15 : Oligonucléotides utilisés pour la construction de la bibliothèque d'aptamères.**

| | |
|---|---|
| Oligo 5' | **TAATACGACTCACTATAG**GGAGACAAGAATAAACGCTCAA |
| Matrice | GCCTGTTGTGAGCCTCCTGTCGAANNNNNNNNNNNNNNNNNNNNNNNNN TTGAGCGTTTATTCTTGTCTCCC |
| Oligo 3' | GCCTGTTGTGAGCCTCCTGTCGAA |

En bleu, la séquence du promoteur de la polymérase T7.

Cette bibliothèque d'ADN a été utilisée pour transcrire *in vitro* la bibliothèque d'aptamères : 100 µg d'ADN ont été transcrits dans 2 mL de milieu de transcription (20 mM Tris pH 8, 11 mM $MgCL_2$, 2 mM de chaque nucléotide, 2,5 mM DTT, 0,5 mM spermidine) en présence de 10 µL de T7 polymérase, à 37ºC pendant 3 h.

Grace à l'introduction des 4 nucléotides aux 25 positions centrales de la matrice, la complexité de la bibliothèque d'aptamères est théoriquement de $4^{25}$, soit 1,12 x $10^{15}$. Or, le rendement de cette transcription initiale de la bibliothèque, après purification sur gel est de 90 µg d'ARN, c'est à dire, 2,24 x $10^{15}$ molécules d'ARN. Donc, on peut espérer que en l'absence de biais au cours de la synthèse des oligonucléotides et de la transcription, 2 ml de milieu reactionnel permet à chaque molécule d'être représentée au moins une fois.

*c. Cycles de sélection*

J'ai utilisé le domaine C-terminal de *Pf*tRip (*Pf*tRip$_{214-402}$) fusionné à la protéine GST comme cible de la sélection. Cinq étapes de sélections successives ont été effectuées (Tableau 16) : La première étape consiste en une sélection essentiellement positive permettant d'enrichir la bibliothèque en aptamères qui se fixent sur la protéine (et accessoirement sur la résine). Toutes les autres étapes de sélection englobent une sélection négative en présence de différents compétiteurs pour éliminer les aptamères qui se fixent : sur la résine (sélection 2), sur la GST (sélections 3 et 4) ou en dehors du site de liaison à l'ARN de *Pf*tRip (sélection 5).

**Tableau 16 : Conditions expérimentales utilisées au cours des différentes étapes de sélection**

| Etapes de sélection | | Elément de sélection | Température | Temps |
|---|---|---|---|---|
| 1ère | Négative | - | - | - |
| | Positive | GST *Pf*tRip$_{214-402}$ | 4ºC | 210 min |
| 2ème | Négative | Agarose de glutathion | 26ºC | 10 min |
| | Positive | GST *Pf*tRip$_{214-402}$ | 26ºC | 30 min |
| 3ème | Négative | GST | 26ºC | 30 min |
| | Positive | GST *Pf*tRip$_{214-402}$ | 26ºC | 30 min |
| 4ème | Négative | GST | 37ºC | 5 min |
| | Positive | GST *Pf*tRip$_{214-402}$ | 37ºC | 5 min |
| 5ème | Négative | GST *Pf*tRip$_{214-402}$ Mut | 26ºC | 30 min |
| | Positive | GST *Pf*tRip$_{214-402}$ | 37ºC | 5 min |

Pour cette expérience, 320 pmol de protéine (GST *Pf*tRip$_{214-402}$, GST ou GST *Pf*tRip$_{214-402}$ Mut) sont immobilisées sur 200 µL d'agarose glutathion (incubation de 15 min à température ambiante, sous agitation). L'excès de tampon est enlevé par centrifugation à température ambiante de 2 min à 5000 rpm (5 µL sont mis sur gel PAGE-SDS pour confirmer l'absence de protéine dans le surnageant).

La résine est incubée dans différentes conditions de température et de temps (Tableau 15) avec la bibliothèque d'aptamères dans un volume total de 1 mL de tampon tampon HEPES-KCl (50 mM HEPES-KOH pH 7,5, 150 mM KCl).

Sélections positives : les aptamères d'intérêt sont ceux qui restent fixés sur la colonne d'affinité. Ils sont lavés avec 5 ml de tampon HEPES-KCl et élués en présence de 400 µL de glutathion réduit 10 mM, 5 min dans la glace.

Sélections négatives : les aptamères d'intérêt sont présents dans la fraction d'aptamères qui ne se lie pas à la colonne.

Les élutions sont traitées par extraction phénolique, et les aptamères sélectionnés sont précipitées à l'éthanol.

Entre chaque cycle de sélection les aptamères sélectionnés sont amplifiés : les culots d'aptamères sont repris dans 50 µL d'eau et sont hybridés 5 min à 70ºC suivis d'une incubation de 10 min dans la glace en présence de 250 pmol d'oligo 3'. Les ARN sont reverse

transcrits en présence de 75 mM de dNTP, 20 unités de AMV-Reverse Transcriptase et du tampon AMV-RT (Life Sciences) dans un volume final de 75 µL pendant 60 min à 42ºC. La réaction est arrêtée par extraction phénolique et précipitation. L'ADN issu de cette réaction est ensuite amplifié par 25 cycles de PCR (1 min à 95ºC, 1 min à 58ºC, 30 sec à 72ºC) dans le milieu indiqué précédemment, et transcrit.

*d. Clonage des candidats*

L'ADN correspondant aux aptamères issus des étapes 4 et 5 de la sélection ont étés clonés dans le plasmide pDrive (Qiagen PCR cloning kit) et transformés dans des cellules Top10. 46 clones ont été séquencés.

# REFERENCES

# A

Achsel, T., and Gross, H.J. (1993). Identity determinants of human tRNA(Ser): sequence elements necessary for serylation and maturation of a tRNA with a long extra arm. The EMBO journal *12*, 3333-3338.

Adhya, S., Ghosh, T., Das, A., Bera, S.K., and Mahapatra, S. (1997). Role of an RNA-binding protein in import of tRNA into Leishmania mitochondria. J Biol Chem *272*, 21396-21402.

Alexandrov, A., Chernyakov, I., Gu, W., Hiley, S.L., Hughes, T.R., Grayhack, E.J., and Phizicky, E.M. (2006). Rapid tRNA decay can result from lack of nonessential modifications. Mol Cell *21*, 87-96.

Amino, R., Giovannini, D., Thiberge, S., Gueirard, P., Boisson, B., Dubremetz, J.F., Prevost, M.C., Ishino, T., Yuda, M., and Menard, R. (2008). Host cell traversal is important for progression of the malaria parasite through the dermis to the liver. Cell host & microbe *3*, 88-96.

Anderson, J., Phan, L., Cuesta, R., Carlson, B.A., Pak, M., Asano, K., Björk, G.R., Tamame, M., and Hinnebusch, A.G. (1998). The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. Genes & development *12*, 3650-3662.

Arisue, N., and Hashimoto, T. (2015). Phylogeny and evolution of apicoplasts and apicomplexan parasites. Parasitology international *64*, 254-259.

Arnason, U., Gullberg, A., and Janke, A. (1997). Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between Xenarthra (Edentata) and Ferungulates. Molecular biology and evolution *14*, 762-768.

Arnez, J.G., Harris, D.C., Mitschler, A., Rees, B., Francklyn, C.S., and Moras, D. (1995). Crystal structure of histidyl-tRNA synthetase from Escherichia coli complexed with histidyl-adenylate. The EMBO journal *14*, 4143-4155.

# B

Bannister, L.H., Hopkins, J.M., Fowler, R.E., Krishna, S., and Mitchell, G.H. (2000). A brief illustrated guide to the ultrastructure of Plasmodium falciparum asexual blood stages. Parasitology today (Personal ed) *16*, 427-433.

Bastien, O., Lespinats, S., Roy, S., Métayer, K., Fertil, B., Codani, J.J., and Maréchal, E. (2004). Analysis of the compositional biases in Plasmodium falciparum genome and proteome using Arabidopsis thaliana as a reference. Gene *336*, 163-173.

Betat, H., Rammelt, C., and Morl, M. (2010). tRNA nucleotidyltransferases: ancient catalysts with an unusual mechanism of polymerization. Cellular and molecular life sciences : CMLS *67*, 1447-1463.

Begley, U., Dyavaiah, M., Patil, A., Rooney, J.P., DiRenzo, D., Young, C.M., Conklin, D.S., Zitomer, R.S., and Begley, T.J. (2007). Trm9-catalyzed tRNA modifications link translation to the DNA damage response. Mol Cell *28*, 860-870.

Bhardwaj, N., and Saxena, R.K. (2013). Heterogeneity of reticulocyte population in mouse peripheral blood.

Billker, O., Shaw, M.K., Margos, G., and Sinden, R.E. (1997). The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of Plasmodium berghei in vitro. Parasitology *115 ( Pt 1)*, 1-7.

Bombardier, J.P., and Munson, M. (2015). Three steps forward, two steps back: mechanistic insights into the assembly and disassembly of the SNARE complex. Curr Opin Chem Biol *29*, 66-71.

Bonnefond, L., Giegé, R., and Rudinger-Thirion, J. (2005). Evolution of the tRNA(Tyr)/TyrRS aminoacylation systems. Biochimie *87*, 873-883.

Bour, T., Mahmoudi, N., Kapps, D., Thiberge, S., Bargieri, D., Menard, R., and Frugier, M. (2016). Apicomplexa-specific tRip facilitates import of exogenous tRNAs into malaria parasites. Proc Natl Acad Sci U S A *113*, 4717-4722.

Bouzaidi-Tiali, N., Aeby, E., Charriere, F., Pusnik, M., and Schneider, A. (2007). Elongation factor 1a mediates the specificity of mitochondrial tRNA import in T. brucei. The EMBO journal *26*, 4302-4312.

Bullard, J.M., Cai, Y.C., Demeler, B., and Spremulli, L.L. (1999). Expression and characterization of a human mitochondrial phenylalanyl-tRNA synthetase. J Mol Biol *288*, 567-577.

# C

Calado, A., Treichel, N., Müller, E.C., Otto, A., and Kutay, U. (2002). Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. The EMBO journal *21*, 6216-6224.

Chan, C.T., Pang, Y.L., Deng, W., Babu, I.R., Dyavaiah, M., Begley, T.J., and Dedon, P.C. (2012). Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun *3*, 937.

Chatterjee, K., Majumder, S., Wan, Y., Shah, V., Wu, J., Huang, H.Y., and Hopper, A.K. (2017). Sharing the load: Mex67-Mtr2 cofunctions with Los1 in primary tRNA nuclear export. Genes & development *31*, 2186-2198.

Chatterjee, K., Nostramo, R.T., Wan, Y., and Hopper, A.K. (2018). tRNA dynamics between the nucleus, cytoplasm and mitochondrial surface: Location, location, location. Biochimica et biophysica acta *1861*, 373-386.

Chaudhry, S.R., Lwin, N., Phelan, D., Escalante, A.A., and Battistuzzi, F.U. (2018). Comparative analysis of low complexity regions in Plasmodia. Sci Rep *8*, 335.

Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., Feng, G.H., Peng, H., Zhang, X., Qian, J.*, et al.* (2016). Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science *351*, 397-400.

Cromer, D., Evans, K.J., Schofield, L., and Davenport, M.P. (2006). Preferential invasion of reticulocytes during late-stage Plasmodium berghei infection accounts for reduced circulating reticulocyte levels. Int J Parasitol *36*, 1389-1397.

Cudny, H., and Deutscher, M.P. (1986). High-level overexpression, rapid purification, and properties of Escherichia coli tRNA nucleotidyltransferase. J Biol Chem *261*, 6450-6453.

Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N., and Leberman, R. (1990). A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 A. Nature *347*, 249-255.

Cutts, E.E., Laasch, N., Reiter, D.M., Trenker, R., Slater, L.M., Stansfeld, P.J., and Vakonakis, I. (2017). Structural analysis of P. falciparum KAHRP and PfEMP1 complexes with host erythrocyte spectrin suggests a model for cytoadherent knob protrusions. PLoS pathogens *13*, e1006552.

# D

Davies, H.M., Nofal, S.D., McLaughlin, E.J., and Osborne, A.R. (2017). Repetitive sequences in malaria parasite proteins. FEMS microbiology reviews *41*, 923-940.

Davies, H.M., Thalassinos, K., and Osborne, A.R. (2016). Expansion of Lysine-rich Repeats in Plasmodium Proteins Generates Novel Localization Sequences That Target the Periphery of the Host Erythrocyte. J Biol Chem *291*, 26188-26207.

Delage, L., Dietrich, A., Cosset, A., and Maréchal-Drouard, L. (2003). In vitro import of a nuclearly encoded tRNA into mitochondria of Solanum tuberosum. Mol Cell Biol *23*, 4000-4012.

Delarue, M., and Moras, D. (1993). The aminoacyl-tRNA synthetase family: modules at work. BioEssays : news and reviews in molecular, cellular and developmental biology *15*, 675-687.

Desai, S.A., Krogstad, D.J., and McCleskey, E.W. (1993). A nutrient-permeable channel on the intraerythrocytic malaria parasite. Nature *362*, 643-646.

Dietrich, A., Maréchal-Drouard, L., Carneiro, V., Cosset, A., and Small, I. (1996). A single base change prevents import of cytosolic tRNA(Ala) into mitochondria in transgenic plants. Plant J *10*, 913-918.

Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. PLoS Genet *2*, e221.

Dixon, L.J., Barnes, M., Tang, H., Pritchard, M.T., and Nagy, L.E. (2013). Kupffer cells in the liver. Comprehensive Physiology *3*, 785-797.

Dock, A.C. (1984). Contribution a l'étude structurale de l'ARN de transfert: étude du tRNAAsp de levure (Université Louis Pasteur, Strasbourg).

Dreher, T.W., Uhlenbeck, O.C., and Browning, K.S. (1999). Quantitative assessment of EF-1alpha.GTP binding to aminoacyl-tRNAs, aminoacyl-viral RNA, and tRNA shows close correspondence to the RNA binding properties of EF-Tu. J Biol Chem *274*, 666-672.

Durdevic, Z., Mobin, M.B., Hanna, K., Lyko, F., and Schaefer, M. (2013). The RNA methyltransferase Dnmt2 is required for efficient Dicer-2-dependent siRNA pathway activity in Drosophila. Cell reports *4*, 931-937.

# E

Elliott, D.A., McIntosh, M.T., Hosgood, H.D., Chen, S., Zhang, G., Baevova, P., and Joiner, K.A. (2008). Four distinct pathways of hemoglobin uptake in the malaria parasite Plasmodium falciparum. Proc Natl Acad Sci U S A *105*, 2463-2468.

Emetu, S., Troiano, M., Goldmintz, J., Tomberlin, J., Grelet, S., Howe, P.H., Korey, C., and Geslain, R. (2018). Metabolic Labeling and Profiling of Transfer RNAs Using Macroarrays. J Vis Exp.

Emilsson, V., Näslund, A.K., and Kurland, C.G. (1992). Thiolation of transfer RNA in Escherichia coli varies with growth rate. Nucleic Acids Res *20*, 4499-4505.

Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature *347*, 203-206.

Eriani, G., Karam, J., Jacinto, J., Morris Richard, E., and Geslain, R. (2015). MIST, a Novel Approach to Reveal Hidden Substrate Specificity in Aminoacyl-tRNA Synthetases. PLoS One *10*, e0130042.

# F

Fechter, P., Rudinger, J., Giegé, R., and Théobald-Dietrich, A. (1998). Ribozyme processed tRNA transcripts with unfriendly internal promoter for T7 RNA polymerase: production and activity. FEBS Lett *436*, 99-103.

Ferguson, D.J., Balaban, A.E., Patzewitz, E.M., Wall, R.J., Hopp, C.S., Poulin, B., Mohmmed, A., Malhotra, P., Coppi, A., Sinnis, P.*, et al.* (2014). The repeat region of the circumsporozoite protein is critical for sporozoite formation and maturation in Plasmodium. PLoS One *9*, e113923.

Filisetti, D., Theobald-Dietrich, A., Mahmoudi, N., Rudinger-Thirion, J., Candolfi, E., and Frugier, M. (2013). Aminoacylation of Plasmodium falciparum tRNA(Asn) and insights in the synthesis of asparagine repeats. J Biol Chem *288*, 36361-36371.

Foth, B.J., and McFadden, G.I. (2003). The apicoplast: a plastid in Plasmodium falciparum and other Apicomplexan parasites. International review of cytology *224*, 57-110.

Frevert, U., Galinski, M.R., Hügel, F.U., Allon, N., Schreier, H., Smulevitch, S., Shakibaei, M., and Clavijo, P. (1998). Malaria circumsporozoite protein inhibits protein synthesis in mammalian cells. The EMBO journal *17*, 3816-3826.

Frugier, M., Bour, T., Ayach, M., Santos, M.A., Rudinger-Thirion, J., Theobald-Dietrich, A., and Pizzi, E. (2010). Low Complexity Regions behave as tRNA sponges to help co-translational folding of plasmodial proteins. FEBS Lett *584*, 448-454.

Fu, H., Feng, J., Liu, Q., Sun, F., Tie, Y., Zhu, J., Xing, R., Sun, Z., and Zheng, X. (2009). Stress induces tRNA cleavage by angiogenin in mammalian cells. FEBS Lett *583*, 437-442.

# G

Galli, G., Hofstetter, H., and Birnstiel, M.L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature *294*, 626-631.

Gardner, M.J., Shallom, S.J., Carlton, J.M., Salzberg, S.L., Nene, V., Shoaibi, A., Ciecko, A., Lynn, J., Rizzo, M., Weaver, B.*, et al.* (2002). Sequence of Plasmodium falciparum chromosomes 2, 10, 11 and 14. Nature *419*, 531-534.

Giegé, R. (2008). Toward a more complete view of tRNA biology. Nature structural & molecular biology *15*, 1007-1014.

Giegé, R., Juhling, F., Putz, J., Stadler, P., Sauter, C., and Florentz, C. (2011). Structure of transfer RNAs: similarity and variability. Wiley interdisciplinary reviews RNA *3*, 37-61.

Giegé, R., Sissler, M., and Florentz, C. (1998). Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res *26*, 5017-5035.

Giegé, R., and Springer, M. (2016). Aminoacyl-tRNA Synthetases in the Bacterial World. EcoSal Plus *7*.

Ginsburg, H., Kutner, S., Krugliak, M., and Cabantchik, Z.I. (1985). Characterization of permeation pathways appearing in the host membrane of Plasmodium falciparum infected red blood cells. Molecular and biochemical parasitology *14*, 313-322.

Goodarzi, H., Nguyen, H.C.B., Zhang, S., Dill, B.D., Molina, H., and Tavazoie, S.F. (2016). Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. Cell *165*, 1416-1427.

Goswami, S., Dhar, G., Mukherjee, S., Mahata, B., Chatterjee, S., Home, P., and Adhya, S. (2006). A bifunctional tRNA import receptor from Leishmania mitochondria. Proc Natl Acad Sci U S A *103*, 8354-8359.

Graczyk, D., Ciesla, M., and Boguta, M. (2018). Regulation of tRNA synthesis by the general transcription factors of RNA polymerase III - TFIIIB and TFIIIC, and by the MAF1 protein. Biochimica et biophysica acta *1861*, 320-329.

Guttery, D.S., Roques, M., Holder, A.A., and Tewari, R. (2015). Commit and Transmit: Molecular Players in Plasmodium Sexual Development and Zygote Differentiation. Trends in parasitology *31*, 676-685.

# H

Habtewold, T., Groom, Z., Duchateau, L., and Christophides, G.K. (2015). Detection of viable plasmodium ookinetes in the midguts of anopheles coluzzi using PMA-qrtPCR. Parasites & vectors *8*, 455.

Haiser, H.J., Karginov, F.V., Hannon, G.J., and Elliot, M.A. (2008). Developmentally regulated cleavage of tRNAs in the bacterium Streptomyces coelicolor. Nucleic Acids Res *36*, 732-741.

Hellmuth, K., Lau, D.M., Bischoff, F.R., Künzler, M., Hurt, E., and Simos, G. (1998). Yeast Los1p has properties of an exportin-like nucleocytoplasmic transport factor for tRNA. Mol Cell Biol *18*, 6374-6386.

Helm, M., Giegé, R., and Florentz, C. (1999). A Watson-Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNALys. Biochemistry *38*, 13338-13346.

Hikosaka, K., Kita, K., and Tanabe, K. (2013). Diversity of mitochondrial genome structure in the phylum Apicomplexa. Molecular and biochemical parasitology *188*, 26-33.

Hoagland, M.B., Stephenson, M.L., SCott, J.F., Hecht, L.I., and Zamecnik, P.C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. J Biol Chem *231*, 241-257.

Hofstetter, H., Kressman, A., and Birnstiel, M.L. (1981). A split promoter for a eucaryotic tRNA gene. Cell *24*, 573-585.

Holler, E. (1978). Protein biosynthesis: the codon-specific activation of amino acids. Angew Chem Int Ed Engl *17*, 648-656.

Holzmann, J., Frank, P., Löffler, E., Bennett, K.L., Gerner, C., and Rossmanith, W. (2008). RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. Cell *135*, 462-474.

Hopper, A.K. (2013). Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast Saccharomyces cerevisiae. Genetics *194*, 43-67.

Hopper, A.K., and Huang, H.Y. (2015). Quality Control Pathways for Nucleus-Encoded Eukaryotic tRNA Biosynthesis and Subcellular Trafficking. Mol Cell Biol *35*, 2052-2058.

Hopper, A.K., Pai, D.A., and Engelke, D.R. (2010). Cellular dynamics of tRNAs and their genes. FEBS Lett *584*, 310-317.

Hou, Y.M. (2010). CCA addition to tRNA: implications for tRNA quality control. IUBMB Life *62*, 251-260.

Hou, Y.M., and Schimmel, P. (1988). A simple structural feature is a major determinant of the identity of a transfer RNA. Nature *333*, 140-145.

Huang, H.Y., and Hopper, A.K. (2015). In vivo biochemical analyses reveal distinct roles of beta-importins and eEF1A in tRNA subcellular traffic. Genes & development *29*, 772-783.

Hunter, C.A., Aukerman, M.J., Sun, H., Fokina, M., and Poethig, R.S. (2003). PAUSED encodes the Arabidopsis exportin-t ortholog. Plant Physiol *132*, 2135-2143.

# I

Ibba, M., Bono, J.L., Rosa, P.A., and Söll, D. (1997). Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete Borrelia burgdorferi. Proc Natl Acad Sci U S A *94*, 14383-14388.

Iben, J.R., and Maraia, R.J. (2012). tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. RNA *18*, 1358-1372.

Invergo, B.M., Brochet, M., Yu, L., Choudhary, J., Beltrao, P., and Billker, O. (2017). Sub-minute Phosphoregulation of Cell Cycle Systems during Plasmodium Gamete Formation. Cell reports *21*, 2017-2029.

Ishino, T., Yano, K., Chinzei, Y., and Yuda, M. (2004). Cell-passage activity is required for the malarial parasite to cross the liver sinusoidal cell layer. PLoS biology *2*, E4.

Ivanov, P., Emara, M.M., Villen, J., Gygi, S.P., and Anderson, P. (2011). Angiogenin-induced tRNA fragments inhibit translation initiation. Mol Cell *43*, 613-623.

# J

Jackman, J.E., and Alfonzo, J.D. (2013). Transfer RNA modifications: nature's combinatorial chemistry playground. Wiley interdisciplinary reviews RNA *4*, 35-48.

Jackson, K.E., Habib, S., Frugier, M., Hoen, R., Khan, S., Pham, J.S., Ribas de Pouplana, L., Royo, M., Santos, M.A., Sharma, A.*, et al.* (2011). Protein translation in Plasmodium parasites. Trends in parasitology *27*, 467-476.

Jahn, D., Verkamp, E., and Soll, D. (1992). Glutamyl-transfer RNA: a precursor of heme and chlorophyll biosynthesis. Trends in biochemical sciences *17*, 215-218.

Jakubowski, H., and Goldman, E. (1992). Editing of errors in selection of amino acids for protein synthesis. Microbiol Rev *56*, 412-429.

Janouskovec, J., Horák, A., Oborník, M., Lukes, J., and Keeling, P.J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc Natl Acad Sci U S A *107*, 10949-10954.

Juhling, F., Putz, J., Florentz, C., and Stadler, P.F. (2012). Armless mitochondrial tRNAs in Enoplea (Nematoda). RNA biology *9*, 1161-1166.

# K

Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., Daub, C.O., Kai, C., Kawai, J., Yasuda, J.*, et al.* (2008). Hidden layers of human small RNAs. BMC Genomics *9*, 157.

Kemp, D.J., Coppel, R.L., and Anders, R.F. (1987). Repetitive proteins and genes of malaria. Annual review of microbiology *41*, 181-208.

Kirchner, S., and Ignatova, Z. (2015). Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. Nature reviews Genetics *16*, 98-112.

Kirk, K., and Lehane, A.M. (2014). Membrane transport in the malaria parasite and its host erythrocyte. The Biochemical journal *457*, 1-18.

Koning-Ward, T.F., Gilson, P.R., Boddey, J.A., Rug, M., Smith, B.J., Papenfuss, A.T., Sanders, P.R., Lundie, R.J., Maier, A.G., Cowman, A.F.*, et al.* (2009). A newly discovered protein export machine in malaria parasites. Nature *459*, 945-949.

Kori, L.D., Valecha, N., and Anvikar, A.R. (2018). Insights into the early liver stage biology of Plasmodium. Journal of vector borne diseases *55*, 9-13.

Kuhn, C.D. (2016). RNA versatility governs tRNA function: Why tRNA flexibility is essential beyond the translation cycle. BioEssays : news and reviews in molecular, cellular and developmental biology *38*, 465-473.

Kutay, U., Lipowsky, G., Izaurralde, E., Bischoff, F.R., Schwarzmaier, P., Hartmann, E., and Görlich, D. (1998). Identification of a tRNA-specific nuclear export receptor. Mol Cell *1*, 359-369.

# L

Landfear, S.M. (2011). Nutrient transport and pathogenesis in selected parasitic protozoa. Eukaryot Cell *10*, 483-493.

Levitz, R., Chapman, D., Amitsur, M., Green, R., Snyder, L., and Kaufmann, G. (1990). The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. The EMBO journal *9*, 1383-1389.

Liu, F., Clark, W., Luo, G., Wang, X., Fu, Y., Wei, J., Hao, Z., Dai, Q., Zheng, G., Ma, H*., et al.* (2016). ALKBH1-Mediated tRNA Demethylation Regulates Translation. Cell *167*, 1897.

# M

Machnicka, M.A., Olchowik, A., Grosjean, H., and Bujnicki, J.M. (2014). Distribution and frequencies of post-transcriptional modifications in tRNAs. RNA biology *11*, 1619-1629.

Manzoni, G., Briquet, S., Risco-Castillo, V., Gaultier, C., Topçu, S., Ivănescu, M.L., Franetich, J.F., Hoareau-Coudert, B., Mazier, D., and Silvie, O. (2014). A rapid and robust selection procedure for generating drug-selectable marker-free recombinant malaria parasites. Sci Rep *4*, 4760.

Marks, M., Gupta-Wright, A., Doherty, J.F., Singer, M., and Walker, D. (2014). Managing malaria in the intensive care unit. British journal of anaesthesia *113*, 910-921.

Masaki, H., and Ogawa, T. (2002). The modes of action of colicins E5 and D, and related cytotoxic tRNases. Biochimie *84*, 433-438.

McClain, W.H. (1993). Rules that govern tRNA identity in protein synthesis. J Mol Biol *234*, 257-280.

McFadden, G.I., and Yeh, E. (2017). The apicoplast: now you see it, now you don't. Int J Parasitol *47*, 137-144.

Ménard, R., Tavares, J., Cockburn, I., Markus, M., Zavala, F., and Amino, R. (2013). Looking under the skin: the first steps in malarial infection and immunity. Nature reviews Microbiology *11*, 701-712.

Milani, K.J., Schneider, T.G., and Taraschi, T.F. (2015). Defining the morphology and mechanism of the hemoglobin transport pathway in Plasmodium falciparum-infected erythrocytes. Eukaryot Cell *14*, 415-426.

Milton, M.E., and Nelson, S.W. (2016). Replication and maintenance of the Plasmodium falciparum apicoplast genome. Molecular and biochemical parasitology *208*, 56-64.

Morales, A.J., Swairjo, M.A., and Schimmel, P. (1999). Structure-specific tRNA-binding protein from the extreme thermophile Aquifex aeolicus. The EMBO journal *18*, 3475-3483.

Mota, M.M., and Rodriguez, A. (2001). Migration through host cells by apicomplexan parasites. Microbes and infection *3*, 1123-1128.

Mota, M.M., and Rodriguez, A. (2002). Invasion of mammalian host cells by Plasmodium sporozoites. BioEssays : news and reviews in molecular, cellular and developmental biology *24*, 149-156.

Motorin, Y., and Helm, M. (2010). tRNA stabilization by modified nucleotides. Biochemistry *49*, 4934-4944.

Moura, G.R., Pinheiro, M., Freitas, A., Oliveira, J.L., Frommlet, J.C., Carreto, L., Soares, A.R., Bezerra, A.R., and Santos, M.A. (2015). Correction: Species-Specific Codon Context Rules Unveil Non-Neutrality Effects of Synonymous Mutations. PLoS One *10*, e0145593.

# N

Nisbet, R.E., and McKenzie, J.L. (2016). Transcription of the apicoplast genome. Molecular and biochemical parasitology *210*, 5-9.

Nomanbhoy, T., Morales, A.J., Abraham, A.T., Vortler, C.S., Giegé, R., and Schimmel, P. (2001). Simultaneous binding of two proteins to opposite sides of a single transfer RNA. Nature structural biology *8*, 344-348.

Novoa, E.M., and Ribas de Pouplana, L. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. Trends Genet *28*, 574-581.

Nozawa, K., Ishitani, R., Yoshihisa, T., Sato, M., Arisaka, F., Kanamaru, S., Dohmae, N., Mangroo, D., Senger, B., Becker, H.D.*, et al.* (2013). Crystal structure of Cex1p reveals the mechanism of tRNA trafficking between nucleus and cytoplasm. Nucleic Acids Res *41*, 3901-3914.

# O

Obado, S.O., Glover, L., and Deitsch, K.W. (2016). The nuclear envelope and gene organization in parasitic protozoa: Specializations associated with disease. Molecular and biochemical parasitology *209*, 104-113.

Ohira, T., and Suzuki, T. (2011). Retrograde nuclear import of tRNA precursors is required for modified base biogenesis in yeast. Proc Natl Acad Sci U S A *108*, 10502-10507.

Orellana, O., Cooley, L., and Söll, D. (1986). The additional guanylate at the 5' terminus of Escherichia coli tRNAHis is the result of unusual processing by RNase P. Mol Cell Biol *6*, 525-529.

OMS (2016). Malaria World Report.

OMS (2017). Malaria World Report.

# P

Packard, R.M. (2014). The origins of antimalarial-drug resistance. The New England journal of medicine *371*, 397-399.

Perona, J.J., Rould, M.A., and Steitz, T.A. (1993). Structural basis for transfer RNA aminoacylation by Escherichia coli glutaminyl-tRNA synthetase. Biochemistry *32*, 8758-8771.

Petrov, A., Wu, T., Puglisi, E.V., and Puglisi, J.D. (2013). RNA purification by preparative polyacrylamide gel electrophoresis.

Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. Genes & development *24*, 1832-1860.

Pinheiro, M., Afreixo, V., Moura, G., Freitas, A., Santos, M.A., and Oliveira, J.L. (2006). Statistical, computational and visualization methodologies to unveil gene primary structure features. Methods Inf Med *45*, 163-168.

Pino, P., Aeby, E., Foth, B.J., Sheiner, L., Soldati, T., Schneider, A., and Soldati-Favre, D. (2010). Mitochondrial translation in absence of local tRNA aminoacylation and methionyl tRNA Met formylation in Apicomplexa. Molecular microbiology *76*, 706-718.

Pizzi, E., and Frontali, C. (2001). Low-complexity regions in Plasmodium falciparum proteins. Genome Res *11*, 218-229.

Preiser, P., Williamson, D.H., and Wilson, R.J. (1995). tRNA genes transcribed from the plastid-like DNA of Plasmodium falciparum. Nucleic Acids Res *23*, 4329-4336.

Putz, J., Giegé, R., and Florentz, C. (2010). Diversity and similarity in the tRNA world: overall view and case study on malaria-related tRNAs. FEBS Lett *584*, 350-358.

# R

Rehkopf, D.H., Gillespie, D.E., Harrell, M.I., and Feagin, J.E. (2000). Transcriptional mapping and RNA processing of the Plasmodium falciparum mitochondrial mRNAs. Molecular and biochemical parasitology *105*, 91-103.

Rich, A., and RajBhandary, U.L. (1976). Transfer RNA: molecular structure, sequence, and properties. Annual review of biochemistry *45*, 805-860.

Rinehart, J., Krett, B., Rubio, M.A., Alfonzo, J.D., and Söll, D. (2005). Saccharomyces cerevisiae imports the cytosolic pathway for Gln-tRNA synthesis into the mitochondrion. Genes & development *19*, 583-592.

Risco-Castillo, V., Topcu, S., Marinach, C., Manzoni, G., Bigorgne, A.E., Briquet, S., Baudin, X., Lebrun, M., Dubremetz, J.F., and Silvie, O. (2015). Malaria Sporozoites Traverse Host Cells within Transient Vacuoles. In Cell host & microbe (United States), pp. 593-603.

Roy, H., and Ibba, M. (2008). RNA-dependent lipid remodeling by bacterial multiple peptide resistance factors. Proc Natl Acad Sci U S A *105*, 4667-4672.

Rubio, M.A., and Hopper, A.K. (2011). Transfer RNA travels from the cytoplasm to organelles. Wiley interdisciplinary reviews RNA *2*, 802-817.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., and Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). Science *252*, 1682-1689.

Ruggero, K., Guffanti, A., Corradin, A., Sharma, V.K., De Bellis, G., Corti, G., Grassi, A., Zanovello, P., Bronte, V., Ciminale, V.*, et al.* (2014). Small noncoding RNAs in cells transformed by human T-cell leukemia virus type 1: a role for a tRNA fragment as a primer for reverse transcriptase. Journal of virology *88*, 3612-3622.

# S

Saks, M.E., Sampson, J.R., and Abelson, J.N. (1994). The transfer RNA identity problem: a search for rules. Science *263*, 191-197.

Salinas, T., Duchêne, A.M., Delage, L., Nilsson, S., Glaser, E., Zaepfel, M., and Maréchal-Drouard, L. (2006). The voltage-dependent anion channel, a major component of the

tRNA import machinery in plant mitochondria. Proc Natl Acad Sci U S A *103*, 18362-18367.

Salinas, T., El Farouk-Ameqrane, S., Ubrig, E., Sauter, C., Duchêne, A.M., and Maréchal-Drouard, L. (2014). Molecular basis for the differential interaction of plant mitochondrial VDAC proteins with tRNAs. Nucleic Acids Res *42*, 9937-9948.

Salinas-Giegé, T., Giegé, R., and Giegé, P. (2015). tRNA biology in mitochondria. Int J Mol Sci *16*, 4518-4559.

Schatz, G., and Dobberstein, B. (1996). Common principles of protein translocation across membranes. Science *271*, 1519-1526.

Schelcher, C., Sauter, C., and Giegé, P. (2016). Mechanistic and Structural Studies of Protein-Only RNase P Compared to Ribonucleoproteins Reveal the Two Faces of the Same Enzymatic Activity. Biomolecules *6*.

Schimmel, P. (2017). The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. Nature reviews Molecular cell biology.

Schneider, A. (2011). Mitochondrial tRNA import and its consequences for mitochondrial translation. Annual review of biochemistry *80*, 1033-1053.

Seidman, D., Johnson, D., Gerbasi, V., Golden, D., Orlando, R., and Hajduk, S. (2012). Mitochondrial membrane complex that contains proteins necessary for tRNA import in Trypanosoma brucei. J Biol Chem *287*, 8892-8903.

Senger, B., Auxilien, S., Englisch, U., Cramer, F., and Fasiolo, F. (1997). The modified wobble base inosine in yeast tRNAIle is a positive determinant for aminoacylation by isoleucyl-tRNA synthetase. Biochemistry *36*, 8269-8275.

Seow, F., Sato, S., Janssen, C.S., Riehle, M.O., Mukhopadhyay, A., Phillips, R.S., Wilson, R.J., and Barrett, M.P. (2005). The plastidic DNA replication enzyme complex of Plasmodium falciparum. Molecular and biochemical parasitology *141*, 145-153.

Sharma, A. and Sharma (2015). Plasmodium falciparum mitochondria import tRNAs along with an active phenylalanyl-tRNA synthetase. The Biochemical journal *465*, 459-469.

Smith, D.W., and McNamara, A.L. (1972). The transfer RNA content of rabbit reticulocytes: enumeration of the individual species per cell. Biochimica et biophysica acta *269*, 67-77.

Soma, A., Kumagai, R., Nishikawa, K., and Himeno, H. (1996). The anticodon loop is a major identity determinant of Saccharomyces cerevisiae tRNA(Leu). J Mol Biol *263*, 707-714.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.

Spielmann, T., and Gilberger, T.W. (2009). Protein export in malaria parasites: do multiple export motifs add up to multiple export pathways? Trends in parasitology *26*, 6-10.

Spillman, N.J., Beck, J.R., and Goldberg, D.E. (2015). Protein export into malaria parasite-infected erythrocytes: mechanisms and functional consequences. Annual review of biochemistry *84*, 813-841.

Stoltenburg, R., Reinemann, C., and Strehlitz, B. (2007). SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. Biomol Eng *24*, 381-403.

Sturchler, C., Westhof, E., Carbon, P., and Krol, A. (1993). Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). Nucleic Acids Res *21*, 1073-1079.

Sundaralingam, M., Rubin, J.R., and Cannon, J.F. Nonenzymatic hydrolysis of RNA: Pb(II)-catalyzed site specific hydrolysis of transfer RNA. The role of the tertiary folding of the polynucleotide chain.

# T

Takano, A., Endo, T., and Yoshihisa, T. (2005). tRNA actively shuttles between the nucleus and cytosol in yeast. Science *309*, 140-142.

Takano, A., Kajita, T., Mochizuki, M., Endo, T., and Yoshihisa, T. (2015). Cytosolic Hsp70 and co-chaperones constitute a novel system for tRNA import into the nucleus. eLife *4*.

Tarun, A.S., Peng, X., Dumpit, R.F., Ogata, Y., Silva-Rivera, H., Camargo, N., Daly, T.M., Bergman, L.W., and Kappe, S.H. (2008). A combined transcriptome and proteome survey of malaria parasite liver stages. Proc Natl Acad Sci U S A *105*, 305-310.

Tavares, J., Formaglio, P., Thiberge, S., Mordelet, E., Van Rooijen, N., Medvinsky, A., Menard, R., and Amino, R. (2013). Role of host cell traversal by the malaria sporozoite during liver infection. The Journal of experimental medicine *210*, 905-915.

Thanaraj, T.A., and Argos, P. (1996). Ribosome-mediated translational pause and protein domain organization. Protein Sci *5*, 1594-1612.

Thompson, D.M., and Parker, R. (2009). Stressing out over tRNA cleavage. Cell *138*, 215-219.

Tschopp, F., Charriere, F., and Schneider, A. (2011). In vivo study in Trypanosoma brucei links mitochondrial transfer RNA import to mitochondrial protein import. EMBO Rep *12*, 825-832.

# V

Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J Mol Biol *180*, 549-576.

# W

Wass, M.N., Stanway, R., Blagborough, A.M., Lal, K., Prieto, J.H., Raine, D., Sternberg, M.J., Talman, A.M., Tomley, F., Yates, J.*, et al.* (2012). Proteomic analysis of Plasmodium in the mosquito: progress and pitfalls. Parasitology *139*, 1131-1145.

Watanabe, K., Ijiri, K., and Ohtsuki, T. (2014). mTOR regulates the nucleoplasmic diffusion of Xrn2 under conditions of heat stress. FEBS Lett *588*, 3454-3460.

Weiner, A.M. (2004). tRNA maturation: RNA polymerization without a nucleic acid template. Curr Biol *14*, R883-885.

Widmann, M., Clairo, M., Dippon, J., and Pleiss, J. (2008). Analysis of the distribution of functionally relevant rare codons. BMC Genomics *9*, 207.

Wyatt, J.R., Chastain, M., and Puglisi, J.D. (1991). Synthesis and purification of large amounts of RNA oligonucleotides. Biotechniques *11*, 764-769.

# X

Xiong, Y., and Steitz, T.A. (2006). A story with a good ending: tRNA 3'-end maturation by CCA-adding enzymes. Curr Opin Struct Biol *16*, 12-17.

Xue, H.Y., and Forsdyke, D.R. (2003). Low-complexity segments in Plasmodium falciparum proteins are primarily nucleic acid level adaptations. Molecular and biochemical parasitology *128*, 21-32.

Xue, S., Calvin, K., and Li, H. (2006). RNA recognition and cleavage by a splicing endonuclease. Science *312*, 906-910.

# Y

Yang, A.S.P., O'Neill, M.T., Jennison, C., Lopaticki, S., Allison, C.C., Armistead, J.S., Erickson, S.M., Rogers, K.L., Ellisdon, A.M., Whisstock, J.C.*, et al.* (2017). Cell Traversal Activity Is Important for Plasmodium falciparum Liver Infection in Humanized Mice. Cell reports *18*, 3105-3116.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene *33*, 103-119.

Yokogawa, T., Kitamura, Y., Nakamura, D., Ohno, S., and Nishikawa, K. (2010). Optimization of the hybridization-based method for purification of thermostable tRNAs in the presence of tetraalkylammonium salts. Nucleic Acids Res *38*, e89.

# Z

Zasloff, M. (1983). tRNA transport from the nucleus in a eukaryotic cell: carrier-mediated translocation process. Proc Natl Acad Sci U S A *80*, 6436-6440.

UNIVERSITÉ DE STRASBOURG

Marta Cela

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

# Import des ARNt dans *Plasmodium : Sélection à l'entrée ?*

## Résumé en français

Mon étude a porté sur la spécificité de l'interaction entre deux protéines du parasite du paludisme (Plasmodium), tRip (tRNA import protein) et la tyrosyl-ARNt synthétase apicoplastique (api-TyrRS), avec l'ARN de transfert (ARNt).

Plasmodium est un parasite intracellulaire qui conserve une organelle vestigiale, l'apicoplaste, qui possède son propre système de traduction. J'ai adapté la séquence de l'ARN messager pour produire l'api-TyrRS in vitro, et j'ai étudié la spécificité de la reconnaissance de l'ARNtTyr apicoplastique, qui évite les interactions erronées plutôt que de favoriser les correctes.

La protéine tRip est située à la surface du parasite et est responsable de l'import des ARNt de l'hôte. Mes résultats suggèrent que cet import à lieu pendant la phase sanguine du parasite. Elle ne reconnait pas tous les ARNt de la même façon. Les modifications post-transcriptionnelles modulent l'affinité de tRip, et potentiellement, le taux d'import de cet ARNt. Finalement, j'ai identifié par SELEX une séquence nucléotidique qui se lie spécifiquement à tRip, un début pour la conception d'une molécule qui ciblerait spécifiquement le parasite du paludisme.

Mots-clés : Plasmodium, ARNt, import d'ARNt, aminoacylation

## Résumé en anglais

My study focused on the specificity of the interaction between two proteins of the malaria parasite (Plasmodium), tRip (tRNA import protein) and the apicoplastic tyrosyl-tRNA synthetase (api-TyrRS), with the transfer RNA (tRNA).

Plasmodium is an intracellular parasite with a vestigial organelle, the apicoplast, which has its own translation system. The messenger RNA sequence was adapted to produce api-TyrRS in vitro, and I studied the specificity of apicoplastic tRNATyr recognition, which avoids erroneous interactions rather than favoring the correct ones.

The tRip protein is located on the surface of the parasite, and is responsible for importing tRNAs from the host. My results suggest that this import takes place during the blood phase of the parasite. In addition, not all tRNAs are recognized uniformly. The post-transcriptional modifications of the tRNAs define the affinity of tRip, and potentialy, the import rate of this tRNA. Finally, I identified a short nucleotide sequence that binds specifically to tRip. It is a good starting point for designing a molecule that specifically targets the malaria parasite.

Key words : Plasmodium, tRNA, tRNA import, aminoacylation